UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

DIONATHAN LUAN DE VARGAS

EXTRAÇÃO E SELEÇÃO DE CARACTERÍSTICAS DE SINAIS DE ELETROENCEFALOGRAMA PARA DIAGNÓSTICO DE CRISES EPILÉPTICAS

PATO BRANCO

2022

DIONATHAN LUAN DE VARGAS

EXTRAÇÃO E SELEÇÃO DE CARACTERÍSTICAS DE SINAIS DE ELETROENCEFALOGRAMA PARA DIAGNÓSTICO DE CRISES EPILÉPTICAS

Feature extraction and selection from electroencephalogram signals for epileptic seizure diagnosis

Dissertação apresentado como requisito para obtenção do título de Mestre em Engenharia Elétrica do Programa de Pós-Graduação em Engenharia Elétrica da Universidade Tecnológica Federal do Paraná.

Orientador: Dr. Marcelo Teixeira

Coorientador: Dr. Jefferson Tales Oliva

PATO BRANCO

2022

4.0 Internacional

Esta licença permite compartilhamento, remixe, adaptação e criação a partir do trabalho, mesmo para fins comerciais, desde que sejam atribuídos créditos ao(s) autor(es). Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.

Ministério da Educação
Universidade Tecnológica Federal do Paraná
Campus Pato Branco

DIONATHAN LUAN DE VARGAS

FEATURE EXTRACTION AND SELECTION FROM ELECTROENCEPHALOGRAMS SIGNALS FOR EPILEPTIC SEIZURE DIAGNOSIS

Trabalho de pesquisa de mestrado apresentado como requisito para obtenção do título de Mestre Em Engenharia Elétrica da Universidade Tecnológica Federal do Paraná (UTFPR). Área de concentração: Sistemas E Processamento De Energia.

Data de aprovação: 23 de Setembro de 2022

Dr. Marcelo Teixeira, Doutorado - Universidade Tecnológica Federal do Paraná

Dr. Dalcimar Casanova, Doutorado - Universidade Tecnológica Federal do Paraná

Dr. Joao Luis Garcia Rosa, Doutorado - Universidade de São Paulo (Usp)

Documento gerado pelo Sistema Acadêmico da UTFPR a partir dos dados da Ata de Defesa em 23/09/2022.

Dedico este trabalho ao meu filho e a minha esposa, Priscila, que me acompanhou e me apoiou nessa jornada.

AGRADECIMENTOS

A Deus, por iluminar a minha mente nos momentos difíceis, me dando força e coragem para não desistir.

A minha esposa Priscila que aceitou enfrentar essa jornada de estresse e realizações que é o mestrado e sempre me apiou e comemorou comigo a cada objetivo alcançado. Ao filho que ainda está em seu ventre. Que um dia ele possa ter orgulho do pai que ele tem.

Agradeço a minha mãe Ieda e a minha avó materna Luiza (em memória), que com humildade e honestidade, sempre me incentivaram a estudar e a ser sempre melhor.

A minha cachorrinha Selena, que sempre esteve deitada ao meu lado, enquanto eu estava estudando.

A meus orientadores Dr. Marcelo Teixeira e Jefferson Tales Oliva que me auxiliaram e estiveram presentes sempre que necessitei, contribuindo com o desenvolvimento do trabalho e me inspirando com essa paixão que eles têm pela ciência e pesquisa. Agradeço também ao Dr. Dalcimar Casanova e ao Dr. João Luis Garcia Rosa, membros da banca avaliadora, que contribuiram muito com o trabalho. E assim agradeço também a todos os professores que de alguma forma me ensinaram a base do conhecimento para chegar nessa nível de aprendizado.

Aos colegas de trabalho e amigos Clecio Jung e Darlan Klotz pelos momentos de descontração e por me mostrarem o caminho das pedras dentro do PPGEE-UTFPR.

Por fim, quero agradecer a todos os meus demais amigos, com quem divido todas as minhas alegrias e angústicas, especialmente a: Junior Libardoni, Rogerio Lovato, Ricardo Karpinski e Jorge Ramos que fizeram meus dias mais felizes.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001, do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), da Fundação Araucária (FA) e da Financiadora de Estudos e Projetos (FINEP). Além disso, os autores agradecem à Universidade Tecnológica Federal do Paraná (UTFPR) e ao Instituto Federal Catarinense (IFC) - Campus Luzerna pela oportunidade de desenvolver este trabalho.

"Se cheguei até aqui foi porque me apoiei no
ombro dos gigantes."(Isaac Newton)

RESUMO

A epilepsia é uma das doenças mais comuns e afeta aproximadamente 50 milhões de pessoas em todo o mundo. Seu diagnóstico requer a análise de um eletroencefalograma, que mede a atividade elétrica cerebral representada por séries temporais. A análise dos segmentos do eletroencefalograma depende da interpretação humana, o que pode levar a resultados divergentes, ser tedioso, impreciso e propenso a erros. Além disso, estima-se que mais de 80% dos exames de epilepsia não retornam nenhuma anomalia, desperdiçando o esforço de análise. Este artigo propõe uma maneira automática de analisar segmentos de eletroencefalograma. A abordagem é baseada na combinação de métodos de análise multiespectral, seleção de características e aprendizado de máquina para redução de dimensionalidade na classificação de eletroencefalograma. Como nossa principal contribuição, propusemos uma metodologia para minimizar o número de características necessários para o treinamento do classificador, minimizando a interferência de medidas irrelevantes. As características selecionadas foram utilizadas como entrada de cinco algoritmos de aprendizado de máquina para classificar segmentos de eletroencefalograma. Um conjunto mínimo de características é escolhido para cada método de seleção, e os resultados são comparados selecionando o melhor subconjunto de características. Neste trabalho, características baseadas em estatística, energia, entropia e medidas específicas foram extraídas das quatro representações do eletroencefalograma. Essas características foram selecionados por técnicas de seleção de características e usadas como entrada em algoritmos de aprendizado de máquina. Na avaliação experimental, os resultados comprovam estatisticamente que essa abordagem é tão eficiente quanto usar o conjunto de dados de características completo para a construção do classificador. Os classificadores propostos superam as métricas avaliativas de trabalhos já publicados na literatura, alcançando uma precisão entre 87,2 e 90,99%, além de reduzirem consideravelmente a quantidade de características necessárias, de 285 para 30, mantendo a acurácia do modelo.

Palavras-chave: classificação multiclasse; eletroencefalograma; epilepsia; redução de dimensionalidade; seleção de características.

ABSTRACT

Epilepsy is one of the most common diseases and affects approximately 50 million people worldwide. Its diagnosis requires analyzing an electroencephalogram (EEG), which measures brain electrical activity represented by time series. The analysis of EEG segments depends on human interpretation, which may lead to divergent results, and be tedious, imprecise, and error-prone. Moreover, one estimates more than 80% of the epilepsy exams return no anomalies at all, wasting the effort of analysis. This paper proposes an automatic way to analyze EEG segments. The approach is based on the combination of multispectral analysis, feature selection and machine learning (ML) methods for dimensionality reduction in EEG classification. As our main contribution, we proposed a methodology to minimize the number of features required for classifier training, minimizing the interference of irrelevant measures. The selected features were used as input of five ML algorithms to classify EEG segments. A minimum set of features is chosen for each selection method, and the results are compared select the best feature subset. In this work, features based on statistics, energy, entropy and specific measures were extracted from the four representations of the EEG. These features were selected by feature selection techniques and used as input in ML algorithms. In the experimental evaluation, results statistically prove that this approach is as efficient as using the complete feature dataset for classifier building. The proposed classifiers improve the related literature, reaching an accuracy between 87.2 and 90.99%, and considerably reduce the number of features, from 285 to 30, under the same accuracy scores.

Keywords: dimensionality reduction; electroencephalogram; epilepsy; feature selection; multi-class classification.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

Abreviaturas

Acc.	Accuracy
Pre.	Precision
Sen.	Sensitivity
Spe.	Specificity

Siglas

1DA	1st Order Diagonal Amplitude
2DA	2nd Order Diagonal Amplitude
ALS	Absolute Logarithmic Sum
ANN	Artificial Neural Network
BG	Biespectrogram
CFS	Correlation-based Feature Selection
CM	Confusion Matrix
DT	Decision Tree
EEG	Electroencephalogram
ELU	Exponential Linear Unit
FD	Frequency Domain
FN	False Negatives
FP	False Positives
GED	Graph Eigen Decomposition
HRC	Hit Rate by Class
KNN	K-Nearest Neighbors
MaF1	Macro F1-Score

ML	Machine Learning
MLP	Multi Layer Perceptron
MM	Mean Magnitude
MSE	Mean Squared Error
NB	Naïve Bayes
NE	Normalized Entropy
NLA	Nonlinear Analisys
NPV	Negative Predictive Value
NQE	Normalized Quadratic Entropy
PS	Power Spectrum
Q1	First Quartile
Q2	Second Quartile
Q3	Third Quartile
PPGEE	Programa de Pós Graduação em Engenharia Elétrica
RBF	Radial Base Function
RELU	Rectified Linear Unit
RF	Random Forests
RFE	Recursive Feature Elimination
SBE	Spectral band energy
SBW	Spectral band width
SC	Spectral center
SELU	Scaled Exponential Linear Unit
SG	Spectrogram
STFT	Short-Time Fourier Transform
SVM	Support Vector Machines
TD	Time Domain

TFD	Time-frequency Domain
TN	True Negatives
TP	True Positives
TS	Time Series
UTFPR	Universidade Tecnológica Federal do Paraná
XWC	X-Axis Weighted Center
YWC	Y-Axis Weighted Center

SUMÁRIO

1 INTRODUCTION

Epilepsy is a term used to designate a group of diseases whose main characteristic is the manifestation of excessive electrical activity in nerve cells (MANOLIS *et al.*, 2019). These diseases affect around 50 million people worldwide, of which 80% live in developing countries (World Health Organization, 2019). Of these, about 75% do not have access to adequate treatment, even though in 70% of cases, the disease is controllable with the use of medication or electrical stimulation (World Health Organization, 2014).

Epilepsy is commonly diagnosed through electroencephalography tests, which result from monitoring the brain's electrical activity (TATUM, 2007). This test results in time-series (TS) called electroencephalograms (EEGs) (SHIN *et al.*, 2014). An expert visually analyzes the EEG to detect epileptiform patterns, which are classified into four distinct (FISHER; SCHARFMAN; DECURTIS, 2014; HWANG; GOODMAN; STEVENS, 2019) periods:

- *Ictal*: moment of the epileptic seizure;
- *Post-ictal*: clinical and electroencephalographic manifestations that follow the end of an epileptic seizure;
- *Interictal*: neurological oscillations between the end of the post-ictal period and the beginning of the next crisis. This neurological event is also common in healthy patients, particularly often during sleep;
- *Pre-ictal*: neurological activity immediately before the ictal period.

These neurological periods can still have variations, for example, signals being captured in a hippocampal formation of the brain (TATUM, 2007). So, selecting these signals demands a lot of time and analysis effort, and it is also susceptible to errors of human interpretation and divergence of opinions among specialists on the same abnormality evidenced in EEGs (OLIVA; ROSA, 2019).

Computational approaches emerge as alternatives to assist medical experts in identifying events related to epilepsy in EEGs (LI *et al.*, 2020; VARGAS; OLIVA; TEIXEIRA, 2021; RAMAKRISHNAN; MURUGAVEL, 2019; TÜRK; ÖZERDEM, 2019). However, the implementation of such approaches is considered complex because the EEG segments present aspects of non-stationarity and non-linearity (FREEMAN; QUIROGA, 2012). One of the ways to try to alleviate the difficulty in such studies is by dividing these signals into segments. Also, the extraction of features can aid in tasks such as pattern finding and classifier building. For example, statistical measures (*e.g.* mean and standard deviation) can be computed and used as signal features. This step is also known as feature engineering (HEATON, 2016). These measurements are used as input to machine learning (ML) methods for building models[1] that automate the EEG signal classification process.

[1] In this work, (predictive) models and classifiers are used as synonyms.

Feature extraction is commonly explored in the literature. In signals, features are extracted considering their representations in time domain (TD), frequency domain (FD), time-frequency domain (TFD), and non-linear approach(ACHARYA *et al.*, 2013). However, the combination and relevance of features in the classifier performance is often an imprecise factor. Consequently, this makes it difficult to choose suitable classification models with minimum expected performance (CURA; AKAN, 2021).

As the number of features to be considered increases, so does the computational cost for training and using predictive models for classification. Therefore, too many features can also negatively affect the performance of classifiers. This problem is addressed in the literature as the curse of dimensionality (BELLMAN, 1966).

In the feature set, preprocessing techniques can be applied to improve the classifier performance, such as feature selection, which selects a representative feature subset(KHALID; KHALIL; NASREEN, 2014). By removing elements considered irrelevant, these techniques can positively affect the performance of the ML algorithms for classifier building (WEI *et al.*, 2020; PRASETIYOWATI; MAULIDEVI; SURENDRO, 2020).

In the literature, several feature selection techniques are addressed in processes related to ML for classifying EEG segments (GHAYAB *et al.*, 2016; LI *et al.*, 2016; MURSALIN *et al.*, 2017; MEHLA *et al.*, 2021). However, these works do not make a comparison among these techniques.

Thus, this work presents an approach to finding the smallest subset of features of the respective original set needed to train accurate classifiers. For this, this work proposes to extract a set of characteristics of the four main representations of EEGs explored in the literature. This set of features can be too large for the application of algorithms, causing model overfitting. To improve the classifiers, feature selection techniques are applied and compared to predict the influence that features have on each other and on the final result.

The remaining of this paper is organized as follows: In Section 2 some EEG processing methods developed in related works are presented; in Section **??** presents signal processing methods for signal conversion and feature extraction, ML methods for classifier construction and evaluation techniques are also presented; in Section 4 presents the proposed method and Section 5 presents the final considerations, main contributions of this work and proposals for continuation.

1.1 Objectives

1.1.1 General objective

Select most representative feature sets, based on feature selection techniques, for classification EEG segments related to epileptic seizures.

1.1.2 Specific objectives

- Investigate the relationship that the characteristics have with each other;
- Check which machine learning algorithms can result in building more efficient models;
- Assess whether feature selection methods affect the performance of models.

2 RELATED WORKS

Some works applied for EEG classification to detect epilepsy activities are briefly described in this section. It is important to emphasize that not all related studies used feature selection methods in their experimental evaluation. Works that apply feature selection such as such as Molla *et al.* (2020) and Mehla *et al.* (2021) are found in the literature, however, this step is treated as part of the development of the classification model. The authors do not show the impact that the feature selection implies in the final result because they do not compare models without a selection of features and verify if there was a gain in the results.

In Mursalin *et al.* (2017), classifiers were trained to differentiate ictal activity from other EEG segments. For this, TD and FD features are selected through improved correlation. For classifier building, the Random Forests (RF) method was used, which reached an accuracy of 98.45%.

In Ramakrishnan e Murugavel (2019) models were built to classify five classes of EEGs using features extracted from FD and nonlinear analysis (NLA) in different sub-bands of EEGs signals. The authors introduced fuzzy rules for selecting features and models based on SVM to build classifiers. As a result, they obtained 95% accuracy.

Oliva e Rosa (2019) proposed an intelligent report generation method based on ML using characteristics of TD, FD, TFD, and NLA to classify five EEG classes. In all, a total of 75 different classifiers were built. As a result, a model based on artificial neural networks (ANN) obtained the highest accuracy, which was 86%.

Molla *et al.* (2020) proposed classifiers that separated normal, interictal and ictal segments. TD, FD, and NLA features were extracted in an algorithm based on Graph Eigen Decomposition (GED), which selects the most relevant measures. Finally, ANN was used as a classifier, which obtained an accuracy of 99.55% for binary classification.

The study of Mehla *et al.* (2021) was based on Fourier decomposition methods for non-stationary EEG segments. The Kruskal-Wallis test selected the extracted features. Finally, the selected features were used to train SVM classifiers, obtaining the maximum accuracy of 99.96% for binary classification.

In the literature, approaches based on deep learning are still used, which generally use raw data directly for feature learning and classifier training. Although this approach determines the characteristics automatically, these methods are not dealt with in the present work because they require large databases for the classifier training, and their interpretation is considered complex compared to other types of models (KOHLI *et al.*, 2016; MARCUS, 2020), for example, decision trees (DT).

Table 1 presents comparisons among proposed methods for classifying EEGs and this work. These comparisons were made considering the following criteria:

- *Feature domain:* indicates whether features were extracted from time domain (TD), frequency domain (FD), time-frequency domain (TFD), and non-linear analysis (NLA);
- *Feature selection:* indicates whether any feature selection technique was used;
- *Classification Algorithms:* determines which methods were applied to build the predictive models;
- *Number of classes:* maximum amount of classes explored in the article;
- *Evaluation metrics:* measures used to evaluate predictive models;
- *Cross validation:* indicates which cross-validation approach was used;
- *Statistical tests:* indicates whether statistical tests were performed to analyze the results.

Tabela 1 – Comparison among methods published in related literature

Reference	Domain of features	Features selection	Classification methods	Amount of classes	Metrics evaluated	Cross validation	Statistical tests
Ghayab *et al.* (2016)	TD	Y	SVM	2	Accuracy, Sensitivity, Specificity	N	-
Li *et al.* (2016)	TD, FD, TFD, NLA	Y	SVM	2	Accuracy, Sensitivity, Specificity	2-fold	N
Riaz *et al.* (2016)	TFD	N	SVM	3	Accuracy	10-fold	Y
Mursalin *et al.* (2017)	TD, FD	Y	RF	2	Accuracy, Sensitivity, Specificity	10-fold	N
Oliva e Rosa (2017)	TD	N	DT, 1NN, MLP	2	Error rates, Precision, NPV, Sensitivity, Specificity	10-fold	Y
Hernández *et al.* (2018)	TD, FD, TFD	Y	NB, DT, KNN, SVM, MLP, RF	4	Accuracy	10-fold	N
Sharma, Bhurane e Rajendra Acharya (2018)	TD	N	SVM	3	Accuracy, Sensitivity, Specificity	10-fold	N
Tsipouras (2019)	FD, TFD	N	RF	5	Accuracy	10-fold	N
Hassan *et al.* (2019)	TD, FD	N	ANN	2	Accuracy	5-fold	N

Continued on next page

Tabela 1 – *Continued from previous page*

Reference	Domain of features	Features selection	Classification methods	Amount of classes	Metrics evaluated	Cross validation	Statistical tests
Ramakrishnan e Murugavel (2019)	FD, NLA	Y	SVM	5	Accuracy, Sensitivity, Specificity	20-fold	N
Oliva e Rosa (2019)	TD, FD, TFD, NLA	N	DT, 1NN, NB, ANN, SVM	5	Error rates	10-fold	Y
Raghu *et al.* (2019)	TD	N	SVM, KNN, MLP	3	Accuracy, Sensitivity, Specificity, Precision, NPV	10-fold	N
Molla *et al.* (2020)	TD, FD, NLA	Y	ANN	2	Accuracy, Sensitivity, Specificity	5-fold	N
Qaisar e Hussain (2021)	TFD	Y	DT, KNN, SVM, RF	3	MAP, NMI, F1, Kappa, Specificity	10-fold	N
Vargas, Oliva e Teixeira (2021)	FD	N	ANN	5	Accuracy, Sensitivity, Specificity, F1	10-fold	N

Continued on next page

Tabela 1 – *Continued from previous page*

Reference	Domain of features	Features selection	Classification methods	Amount of classes	Metrics evaluated	Cross validation	Statistical tests
Mehla *et al.* (2021)	FD, TFD	Y	SVM	2	Accuracy, Sensitivity, Specificity, F1, Kappa	10-fold	N
This work	**TD, FD, TFD, NLA**	**Y**	**NB, KNN, SVM, ANN, RF**	**5**	**Accuracy, Hit Rate by Class, Sensitivity, Specificity, Precision, Macro F1**	**10-fold**	**Y**

In this work, we combine features extracted from four different domains (TD, FD, TFD, and NLA). The features were subjected to attribute selection techniques, contributing to accurate classifier building. Evaluative metrics and statistical tests prove that these models can accurately classify five types of EEG segments.

3 PROPOSAL FOR CLASSIFICATION OF EEG SEGMENTS BY MULTIDOMAIN ANALYSIS

An overview of the process of elaborating models for classifying segments of EEGs aimed at diagnosing epilepsy can be seen in Figure 1. The solid lines expose topics explored in this work, while the dashed lines represent options from the literature not emphasized here.

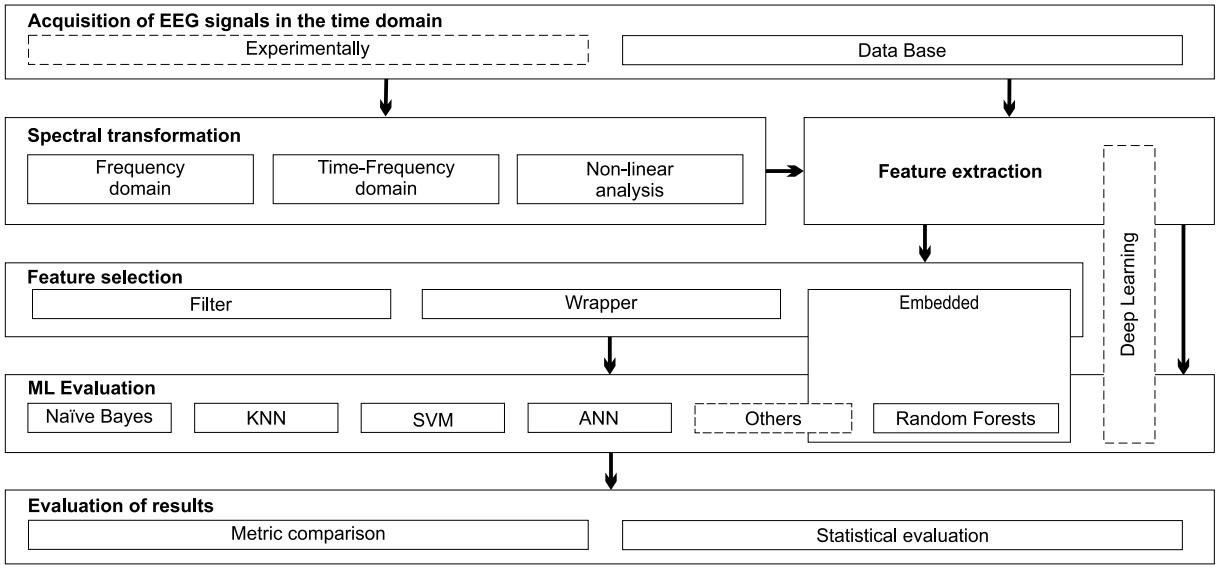


Figura 1 – Proposed methodology for the classification of EEG segments aimed at epilepsy

As shown in Figure 1, EEG segment classification problems are commonly divided into six steps:

3.1 Representations of EEGs

The EEG segments may contain difficult or impossible standards to observe in TD. For example, interictal events may have similar patterns to those found in normal EEG in this domain (TATUM, 2007). In this context, analyzing the signal in other domains, such as in FD, can be more informative (BABADI; BROWN, 2014), as specific frequencies manifest more in some periods of EEGs.

3.1.1 Frequency Domain

For the transposition of time domain observations to the frequency domain, firstly, the signals in TD must be converted using mathematical methods, such as the Fourier (BRIGHAM, 1988) transform. However, the conventional Fourier transform imposes limitations that skewed spectral components can be generated and may affect the reliability of representation (PERCIVAL; WALDEN *et al.*, 1993).

In this sense, the Multitaper method (THOMSON, 1982) was proposed as an improvement to Fourier's transform. This method can be computed by Equation (1), in which $h_{i,k}$ is the

kth eigenvector used as a window function (tapering), $x[k]$ is the k-th frequency component, nth is the length of the signal, j is the imaginary unit of the complex expression, and ω is the angular frequency.

$$X[k] = \sum_{i=1}^{N} h_{i,k} x[n] e^{-j\omega n}. \tag{1}$$

This method consists of the cross-correlation operation between a signal, represented in the TD, and the Euler equation, including k-tapers (BABADI; BROWN, 2014; OLIVA; ROSA, 2021), which are generated by Slepian sequences (SLEPIAN; POLLAK, 1961). These sequences are obtained using Equation (2) where, $0 < |f| < t/2$ is the frequency band analyzed, t is the frequency rate of signal and $0 < \lambda(f) < 1$ (SLEPIAN; POLLAK, 1961).

$$\lambda(f) = \frac{\int_{-f}^{f} |X[k]|^2 df}{\int_{-T/2}^{T/2} |X[k]|^2 df}. \tag{2}$$

The tapers are orthogonal with each other and, therefore, k samples, called subspectrums, are generated and averaged, resulting in one-dimensional FD components. The results generated by the multitaper method are represented as complex numbers. To simplify the components, they can be computed by means of Equation (3) to obtain the Power Spectrum (PS) (KRAMER; GERHARDT, 2012).

$$PS(k) = |X[k]|^2 \tag{3}$$

3.1.2 Time-Frequency Domain

EEGs have non-stationary behavior with respect to frequency components present in the signal that vary over time (MARTÍNEZ-VARGAS; GODINO-LLORENTE; CASTELLANOS-DOMINGUEZ, 2012). Therefore, separate analyses in time and frequency domains may not provide complete signal information. For example, if you wish to observe which frequencies are most incident at a given moment, the information in TD and FD would be insufficient. Thereby, from an adaptation in the Short Time Fourier Transform (STFT), it is possible to use the multitaper method to transform a time domain signal into a time-frequency domain representation (OLIVA *et al.*, 2016).

In the time-frequency domain, the signals are analyzed in time and frequency simultaneously (COHEN, 1995). In this domain, the spectrogram (SG) is generated, which is the dynamic representation of the energy spectral density.

Commonly, the SG can be computed from a STFT that maps a one-dimensional real function $f(t)$ into a complex function $F(k,\omega)$ defined in a two-dimensional space (POULARIKAS, 2010). When related to the SG of EEG segments, considering t time, and k frequency.

The multitaper method can be adapted in Equation (1) to generate SG. In this adaptation, defined by Equation (4), time (t), the window length (w) and the number of tapers of data (q) used for processing signal must be considered (OLIVA, 2019).

$$SG_{t,k} = \left| \frac{1}{q} \sum_{l=0}^{q-1} \sum_{i=t}^{t+w-1} h_{i-k+1,l} X_n e^{-j\omega \frac{i}{n}} \right|^2 . \quad (4)$$

For SG generation, the adapted multitaper method applies the data tapers with length w as sliding windows, while STFT uses a window function in signals (OLIVA, 2019). In Figure 2, a SG example is presented.

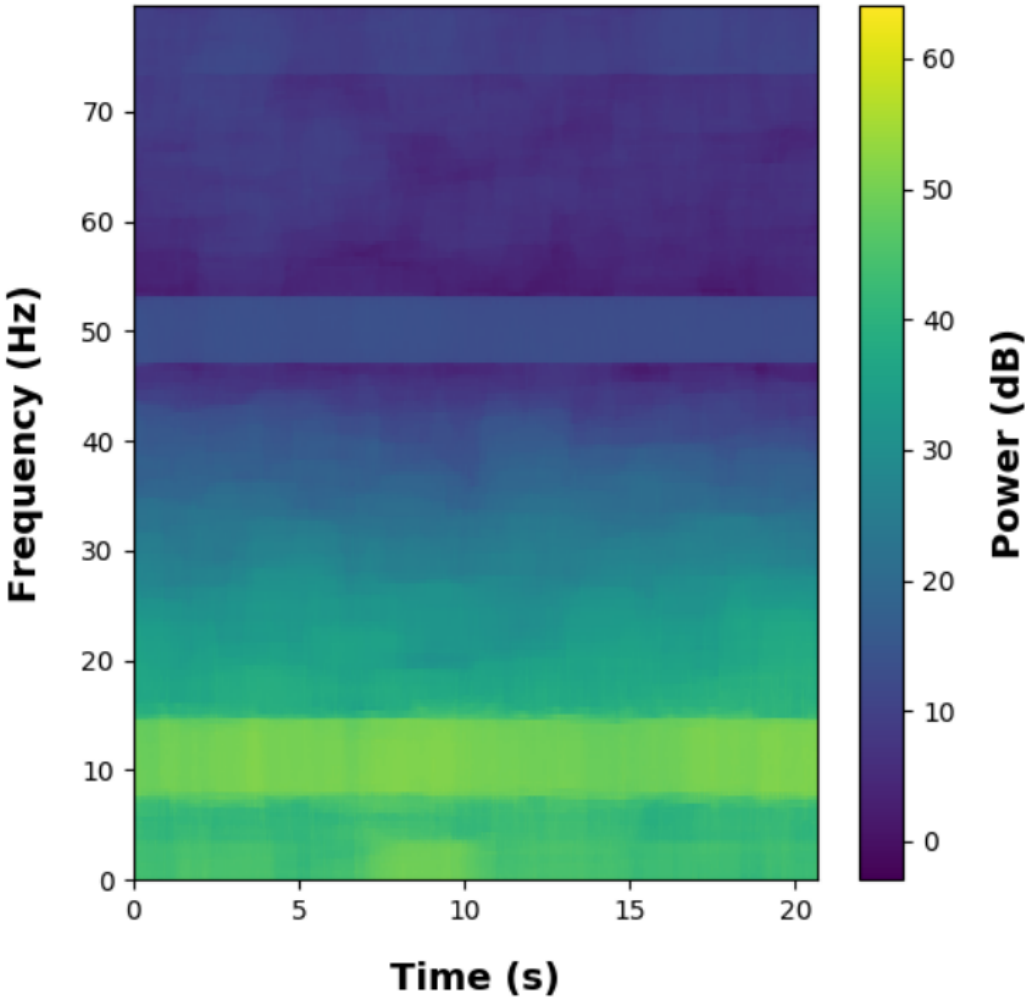


Figura 2 – SG example of a normal signal from the Bonn database (OLIVA; ROSA, 2020).

3.1.3 Non-linear Analysis

Some works related to EEG classification for epilepsy detection address the data representation in the form of nonlinear analysis (LI *et al.*, 2007; RAMAKRISHNAN; MURUGAVEL, 2019; OLIVA; ROSA, 2019; MOLLA *et al.*, 2020) aiming to minimize computational problems

related to this kind of signal, such as non-stationarity and non-linearity. Several nonlinear approaches can be applied to EEG segments, for example, those based on higher order spectrum analysis that uses higher-order statistical measurements for spectral representations (NIKIAS; MENDEL, 1993).

The bispectrum (B) is a third-order statistic commonly applied for EEG analysis. Equation (5) presents a way to obtain bispectrum, where $f1$ and $f2$ are frequency variables and X_f is the result of the multitaper application for the frequency f.

$$B_{f1,f2} = X_{f1} X_{f2} X^*_{f1+f2}. \tag{5}$$

From a bispectrum, the bispectrogram (BG) can be obtained by Equation (6), where $M_{f1,f2}$ is the magnitude measure that represents the relationship between component frequencies $f1$ and $f2$. In this way, BG measures frequency dependency on a signal. Figure 3 represents an example of BG extracted from a normal signal from the Bonn database, where lower frequencies have more dependency than higher frequencies.

$$M_{f1,f2} = |B_{f1,f2}|. \tag{6}$$

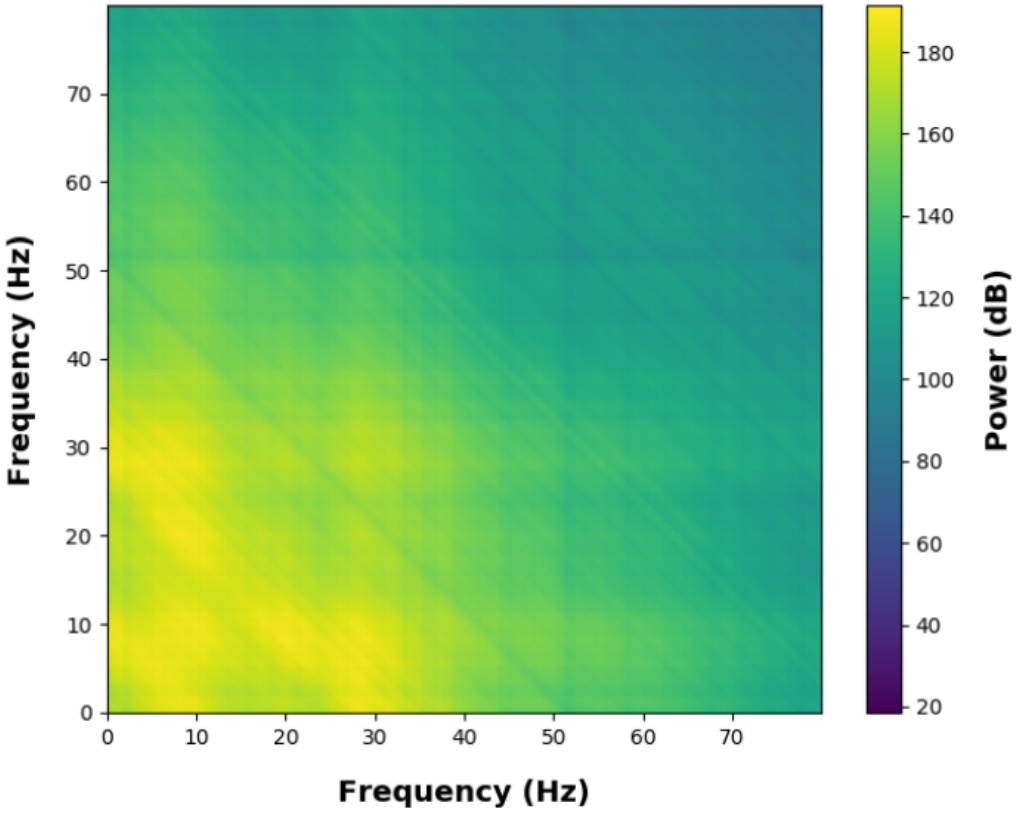


Figura 3 – BG of a normal EEG segment from Bonn base (OLIVA; ROSA, 2020).

The representation of EEGs in the forms presented in Section 3.1 can be used for feature extraction, which is applied as input of ML algorithms. In Section 3.2, some measures computed from these representations are presented.

3.2 Features Extraction

Feature extraction transforms raw data into numerical values by calculating measurements, which can preserve the information of the original data. Generally, feature extraction produces better results than applying the ML algorithm directly to raw database (PAGE *et al.*, 2014).

Several approaches for feature extraction from EEG segments and their representations are proposed. For example, statistical metrics can be computed from TS, PS, and SG, aiming to represent EEG patterns. Commonly computed measures are average, variance, standard deviation, and momentum measurements such as mode, median, asymmetry, and kurtosis (FREEDMAN ROBERT PISANI, 2007).

The minimum and maximum values are also used to quantify the data range and the amplitude defined by the difference between these two metrics. Another statistical parameter includes the coefficient of variation, defined as the ratio between standard deviation and the sample average, representing data dispersion compared to the population average (ZHOU; GAN; SEPULVEDA, 2008).

The first (Q1), second (Q2), and third quartile (Q3) quantify the density of the data. The interquartile interval measures the deviation between the Q1 and Q3. Another measure commonly obtained in EEGs is the line length, which reflects the waveform dimensionality changes, sensitive to signal amplitude variation and frequency (FREEDMAN ROBERT PISANI, 2007).

Energy-based features can also be extracted from TS, PS and SG, such as nonlinear energy, average quadratic value, and crest value (KEETON, 2015).

Other measures commonly computed in TD and FD are entropies. These measures are related to the degree of signal disorder. In other words, the greater the signal disorder, the greater the entropy (SHANNON, 1948). Some entropies are used in the literature as features for EEGs classification models, including the following:

- *Shannon Entropy:* it is used to estimate data distribution and is computed by Equation (7) (SHANNON, 1948), where X represents a data vector defined for x_1, x_2, ..., x_n;

$$\text{Shannon Entropy} = \sum_{i=1}^{n} X_i * \log X_i. \tag{7}$$

- *Rényi Entropy*: it is a generalization of Shannon Entropy to measure data entanglement (RAMALINGAM; KRISHNAN, 2005). This feature can be defined by Equation (8), where

$0 \leq \alpha$ e $\alpha \neq 1$;

$$\text{Rényi Entropy} = \frac{1}{1-\alpha} * \log \sum_{i=1}^{n} (X_i)^{\alpha}. \tag{8}$$

- *Approximate Entropy*: it is a technique used to quantify the regularity and unpredictability of fluctuations on data series (PINCUS; GLADSTONE; EHRENKRANZ, 1991). Approximate Entropy is defined by Equation (9), where $Ci^m(r)$ measures, for a tolerance r, the regularity or frequency of the occurrence of standards similar to a certain standard of length m, concerning a series of data of size N.

$$\begin{aligned} &\text{Approximate Entropy} = \Phi^m(r) - \Phi^{m+1}(r), \\ &\Phi^m(r) = \frac{1}{N-m-1} * \sum_{i=1}^{N-m+1} \log Ci^m(r). \end{aligned} \tag{9}$$

Additionally, the Hurst exponent (QIAN; RASHEED, 2004), which is used as a long-term memory measure, is used for EEG representation. Hurst exponent values range from 0 to 1, referring to the autocorrelations of the data series. Briefly, for values between 0 and 0.5, the signal is characterized as anti-persistent, *i.e.* there is a greater probability that a value does not repeat. For Hurst exponent values between 0.5 and 1, the signal is stated persistent, presenting the likelihood of repeating a value. The signal is stated as aleatory when the value is exactly 0.5.

There are still certain measures that can only be used in certain signal representations, such as the Hjorth parameters (HJORTH, 1970). These parameters are indicators of statistical properties used in signal processing in TD. These parameters are named Activity, Mobility, and Complexity and are defined by:

- *Hjorth Activity*: it represents the signal's power, the variation of a time function. This measure can indicate the surface of the PS at FD. Hjorth activity is computed by Equation (10), where $y(t)$ represents the signal.

$$Activity = var(y(t)). \tag{10}$$

- *Hjorth Mobility*: it represents the average frequency or proportion of the standard deviation of the PS. This measure can be defined by Equation (11).

$$Mobility = \sqrt{\frac{var(\frac{dy(t)}{dt})}{var(y(t))}}. \tag{11}$$

- *Hjorth Complexity*: it represents the change in frequency. The parameter compares the similarity of the signal with a pure sine wave, where the value converges to 1 if the

signal is similar. This parameter is defined by Equation (12).

$$Complexity = \frac{Mobility(\frac{dy(t)}{dt})}{Mobility(y(t))}. \tag{12}$$

Additionally, there are other features that can also be extracted from the FD, of which some examples are presented below:

- *Peak Frequency:* frequency of the maximum power.
- *Spectral Planity:* points out that the spectrum has a similar amount of power in all spectral bands (DUBNOV, 2004).
- *Spectral Entropy:* it is a normalized form of Shannon entropy, which uses the amplitude components of the time series for entropy assessment (VAKKURI *et al.*, 2004).
- *Spectral Centroid:* shows where the centre of mass of PS is located. This feature can be computed using Equation (13), where ϕ represents the spectral component.

$$\text{Spectral Centroid} = \frac{\sum_{i=fmin}^{fmax} \phi * PS_i}{\sum_{i=fmin}^{fmax} PS_i}. \tag{13}$$

The spectral centroid (SC) can also be extracted from SG, however, by Equation (14).

$$\text{SC} = \frac{\sum_{i=fmin}^{fmax} \sum_{j=1}^{L} \phi_i SG_{i,j}}{\sum_{j=1}^{L} SG_{i,j}}. \tag{14}$$

From SG, other features can be extracted, such as spectral band energy (SBE) and spectral bandwidth (SBW). The SBE (Equation (15)) indicates the energy of the spectrum. The SBW is the average deviation of SG concerning SC, whose feature can be obtained by Equation (16) , where L and F are SG length in TD and frequency, respectively (HOSSEINZADEH; KRISHNAN, 2007; LIU; WANG; CHEN, 1998).

$$\text{SBE} = \frac{\sum_{i=fmin}^{fmax} \sum_{j=1}^{L} SG_{i,j}}{\sum_{i=i}^{F} \sum_{j=1}^{L} SG_{i,j}}. \tag{15}$$

$$\text{SBW} = \frac{\sum_{i=fmin}^{fmax} \sum_{j=1}^{L} SG_{i,j}}{\sum_{i=fmin}^{fmax} \sum_{j=1}^{L} (\phi_i - SC)^2 SG_{i,j}}. \tag{16}$$

Generally, in EEG analysis, the PS and SG are divided into representative sub-bands, such as Delta (1 - 4Hz), Theta (4 - 8Hz), Alpha (8 - 12Hz), Beta (12 - 30Hz), and Gamma (30 - 60Hz) (FREEMAN; QUIROGA, 2012). This division is made because some periods of EEG segments have a high incidence in different sub-bands (TATUM, 2007; LI, 2016; STUART *et*

al., 2018). The EEG analysis by sub-band can significantly improve classification results and considerably reduce the computational cost because irrelevant component frequencies are not considered. Consequently, the same feature can be extracted six times, once for each sub-band and once for the entire spectrum.

Finally, BG has specific features that can be extracted. In this paper, we are going to cover the following (ZHOU; GAN; SEPULVEDA, 2008; ACHARYA; SREE; SURI, 2011; ZURLINI *et al.*, 2013):

- *First Order Diagonal Amplitude (1DA)*: is the average of the diagonal of the BG that can be calculated by Equation (17), where N_d is the number of elements contained in the diagonal of the BG;

$$1DA = \sum_{i=1} N_d i * \log M_{i,i}. \tag{17}$$

- *Second Order Diagonal Amplitude (2DA)*: is the variance of the diagonal of BG, defined by Equation (18):

$$2DA = \sum_{i=1} N_d (i - 1DA)^2 * \log M_{i,i}. \tag{18}$$

- *Normalized Entropy (NE)*: is the measure that describes the order of regularity of the BG components. It can be computed by Equation (19) where r is the number of elements contained in the non-redundant area of BG, and Ω represents the positions of the components within this area;

$$NE = -\sum_{m=0}^{r-1} p_m \log p_m, p_m = \frac{M_{f1,f2}}{\sum_{i,j \in \Omega} M_{i,j}} \tag{19}$$

- *Normalized Quadratic Entropy (NQE)*: obtained by Equation (20), this feature quantifies the order of regularity of the quadratic components of BG;

$$NQE = -\sum_{m=0}^{r-1} q_m \log q_m, q_m = \frac{M^2_{f1,f2}}{\sum_{i,j \in \Omega} M^2_{i,j}} \tag{20}$$

- *Average Magnitude (MM)*: is the average value of the components of the non-redundant area of BG. This feature can be measured by Equation (21);

$$MM = \frac{1}{r} \sum_{i,j \in \Omega} M_{i,j}. \tag{21}$$

- *Weighted Center of the X Axis (XWC)*: it is the geometric centre of the BG on the X axis. This measure can be obtained using Equation (22);

$$XWC = \frac{\sum_{i,j \in \Omega} i * M_{i,j}}{\sum_{i,j \in \Omega} M_{i,j}}. \tag{22}$$

- *Weighted Center of the Y Axis (YWC)*: is the geometric centre of the BG on the Y axis. This measure can be calculated using Equation (23);

$$YWC = \frac{\sum_{i,j \in \Omega} j * M_{i,j}}{\sum_{i,j \in \Omega} M_{i,j}}. \tag{23}$$

- *Absolute Logarithmic Sum (ALS)*: is the logarithmic sum of BG. It can be calculated by Equation (24)

$$ALS = \sum_{i,j \in \Omega} \log M_{i,j}. \tag{24}$$

3.3 Selection of Features

The quantity of data needed to achieve a better result exponentially impacts the number of required features. However, the performance of classifiers tends to degrade from a certain number of features, even if they are useful. This phenomenon is known as the curse of dimensionality (BELLMAN, 1966). Thereby, feature (attribute) selection techniques can alleviate this problem.

Feature selection techniques are widely explored in the ML area and are used to reduce the number of features for classifier training to improve time and classification performance (GUYON; ELISSEEFF, 2006).

Some ML problems, like the one in this work, have many features for their representation. However, the greater the number of features, the higher the computational cost for training models using ML algorithms. Also, large amounts of features can require a great deal of system memory. Finally, irrelevant measures can negatively affect the performance of models (KUHN; JOHNSON *et al.*, 2013). In this sense, feature selection methods can alleviate these problems. These methods can be grouped into filter, wrapper, and embedded approaches (KOHAVI; JOHN, 1997).

3.3.1 Filter

In this approach, features are selected before classifier building. Its operation aims to filter out irrelevant features according to some criterion, such as correlation measures (JOHN;

KOHAVI; PFLEGER, 1994). These techniques can evaluate each attribute independently of the others, determining the degree of significance between each attribute and the class (YANG; PEDERSEN, 1997).

The term filter derives from the idea that irrelevant features are filtered from the database before applying the classification algorithm (BLUM; LANGLEY, 1997). The filters use information from the training base itself to choose features to be used later.

In several cases, attribute ranking techniques are commonly used to calculate a metric for the attribute in question and use the ones that rank best. However, these techniques are based on correlation, *e.g.* Correlation-based Feature Selection (CFS). The CFS assume that a relevant subset of features contains features with high correlation with the class and low correlation with each other (HALL, 2000).

Thus, CFS evaluates the subset of features considering the individual predictive capacity of each one together with the degree of redundancy among them. Subsets of features highly correlated with the class while having a low correlation with each other are preferentially selected in these cases.

3.3.2 Wrapper

In this approach, the attribute selection process occurs externally to the model-building algorithm. However, this process uses the model to evaluate the selected features at each iteration.

Thus, feature selection techniques applied to the wrapper approach generate a candidate subset of features. This technique runs the ML algorithm with the selected subset and uses the resulting predictive performance of the classifier as an important measure to evaluate the investigated subset of features. This process is repeated for each subset of features until a given stopping criterion is satisfied (FREITAS, 2002).

Recursive Feature Elimination (RFE) is commonly used in wrapper algorithms. The RFE, based on a ML model, exhaustively traverses all possible subsets of features, eliminating each irrelevant attribute until it reaches the stopping criterion, which can be a specific feature amount or a metric to be achieved.

In general, exhaustive search techniques return better results than filter-based methods, as their construction explores all possible combinations of features. On the other hand, this method takes a long time to execute, making it unfeasible for applications in large search spaces since its complexity grows in the order of $n!$, where n represents the total number of features (PHUONG; LIN; ALTMAN, 2005).

As many features are extracted in this work, the RFE method is unfeasible for this study. So, in our experimental evaluation, the exact search is replaced by feature selection based on a genetic algorithm that reduces the wrapper approach's complexity.

The feature selection algorithm based on a genetic algorithm is performed in the following steps (Figure 4) (ROSTAMI; BERAHMAND; FOROUZANDEH, 2020; BABATUNDE *et al.*, 2014):

1. a population is created, where each individual has a random set of features from the extraction of features step;
2. these individuals are evaluated by a ML algorithm. As a result, an evaluation metric for each individual is obtained;
3. verify whether individuals have met any specified stopping criterion. If so, the features of the best individual in the population are used. If this stopping criterion is not met, the individuals with the lowest classification metric are discarded, and the others are produced at random until the population reaches the amount of the initial population. In this reproduction process, some individuals can mutate and receive some attribute that did not belong to the parents, reducing the chance of the algorithm falling into local minima. The generation is incremented, and the evaluation cycle from step two is repeated.

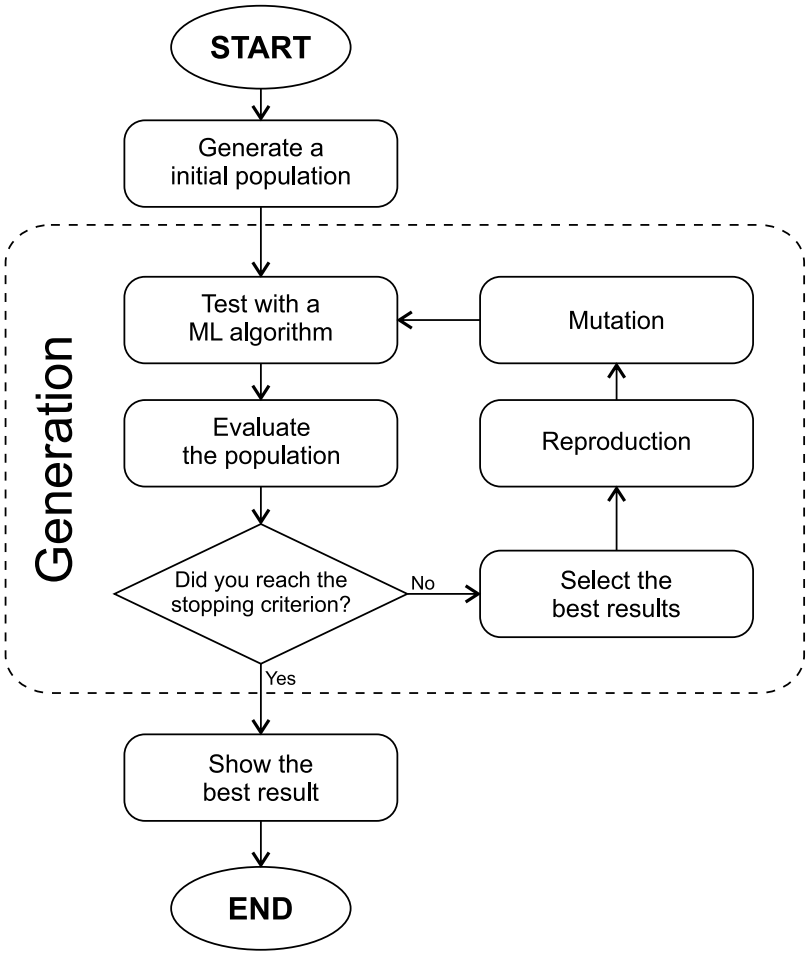


Figura 4 – Flowchart of the operation of the wrapper technique based on genetic algorithm.

3.3.3 Embedded

In this approach, feature selection is performed internally by the ML algorithm itself. Therefore, techniques applied according to the embedded approach select the subset of features during the classifier training(LAL *et al.*, 2006). The main advantage of this approach is that selection works as part of the learning process, which commonly makes them less expensive and allows them to be practicable in high dimensional data. However, as a disadvantage, these approaches are specific to some learning algorithms, for example, those that implement methods based on DT.

3.4 Construction of Classifiers

For the classifier building, selected features are used as input in ML algorithms. Thus, part of the instances represented by these features are used for model training, and the remaining examples are used for the classifier evaluation. How this division is done is discussed in Section 3.5.2.

As seen in Section 2, several classifiers based on ML are used in the literature to classify EEG segments. In this work, we consider some of the commonly used methods, such as Naïve Bayes (NB), K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Artificial Neural Networks (ANN) based on Multi-Layer Perceptron (MLP) and Random Forests (RF) architecture. In this Section, these methods are going to be briefly detailed.

3.4.1 Naïve Bayes

The NB model is a Bayesian learning method that classifies an object into a given class based on the probability that this object belongs to it (MITCHELL *et al.*, 1997). Additionally, this reasoning is based on the assumption that probability distributions govern the quantities of interest and that optimal decisions can be made by reasoning among these probabilities and the observed data(ZHANG, 2004).

The NB algorithm is a direct application of Bayes' theorem, whose fundamental principle describes the probability of the occurrence of an event based on prior knowledge of the conditions that may be related to that fact. This principle is defined by Equation (25), given class y and feature X.

$$P(y|X) = \frac{P(X|y) * P(y)}{P(X)}. \tag{25}$$

The objective is to calculate the posterior probability $P(y|X)$ from the likelihood $P(X|y)$, the prior probability of the hypothesis $P(y)$, and the prior probability that the evidence is true $P(X)$. In other words, the probability that the feature X belongs to the class y depends on the

probability survey in the training step. In the training stage, the probability that samples of class y have presented the feature X, the probability that a class y occurs, and the probability that the feature with the value of X is evaluated (ZHANG, 2004).

3.4.2 K-Nearest Neighbors

KNN is a method based on the concept of distance, that is, on the proximity among the data. The base hypothesis is that similar data tend to be concentrated in the same region in the feature space (LORENA; GAMA; FACELI, 2011). This algorithm classifies a new example according to the proximity of the k-neighbours belonging to the training dataset, where the majority class of the closest k sets is assigned (FIX; HODGES, 1989).

The KNN method requires a low computational effort since the algorithm does not build a model for the data, *i.e.* it only memorizes the elements of the training set. For the classification of new examples, the model calculates the distance of each of them in relation to the test set elements (AHA; KIBLER; ALBERT, 1991).

Different distance measures can be applied to KNN, of which the Euclidean distance is the most common and can be defined by Equation (26), where $d(a,b)$ is the distance between the points a and b, represented by vectors whose elements are the values of each of the n features.

$$d(a,b) = \sqrt{\sum_{1}^{n}(a_i - b_i)^2}. \tag{26}$$

A particular case, which coincides with the simplest version of KNN, is 1-Nearest Neighbor, which assigns the class of the new element to the closest element. However, for modelling certain behaviour, it may be of interest to consider k-neighbors nearest, with k greater than 1, where each neighbour belongs to a class, and the test object is classified according to the majority class. In classification problems, it is usual to use odd or prime numbers greater than two to avoid ties. Also, you can assign a weight to the contribution of each neighbour so that the nearest neighbours have more influence on the decision to which class the element belongs (HASTIE *et al.*, 2009).

3.4.3 Support Vector Machines

This method represents the elements as points in space and finds the best possible separation boundary between classes of a dataset to make them linearly separable. These boundaries are called hyperplanes, and their dimensions are directly proportional to the feature dimension related to the problem (CORTES; VAPNIK, 1995).

Limits are created to choose the best fit for the hyperplane using training data close to the decision boundary, the so-called support vectors. When it is possible to separate the classes,

the algorithm selects the hyperplane with the most significant distance between the support vectors. If it is impossible to do the classification using a hyperplane directly, a kernel function is applied to map the elements in a higher dimension. In this case, the problem would not be linearly separable (LORENA; GAMA; FACELI, 2011; HAYKIN, 2008).

In this way, SVM first separates the elements of each class by a hyperplane and then maximizes the distance between the hyperplane and the closest element. Finally, the test set data is mapped into the same space and predicted to belong to a class based on which side of the hyperplane they are allocated (HEARST, 1998).

3.4.4 Artificial Neural Networks

As a model of ANN, the MLP architecture was used, represented by a parallel distributed system with simple processing units, called artificial neurons, which calculate mathematical functions and are activated or not by this activation function. These neurons are spatially arranged in at least three layers: an input layer, an output layer, and one or more hidden layers. When related to classification, the input layer has several source nodes equal to the number of features of the data set, and the output layer has several neurons equal to the number of classes in the problem (GOODFELLOW; BENGIO; COURVILLE, 2016). These layers are interconnected to each other by weighted connections. The weights of the connections between neurons store the knowledge of the model.

The algorithm backpropagation (RUMELHART; HINTON; WILLIAMS, 1986) is commonly used to train these networks, and it is divided into two stages: forward and backward.

The forward step uses the training set as the network input and propagates it layer by layer until the outputs are produced. These outputs are compared with the true outputs, whose difference is called error.

In the second step, called backward, the error is backpropagated through the network by adjusting the weights for the true classes of the training set.

In short, the successive applications of the forward and backward phases adjust the weights to each interaction, implying a reduction in errors and making the network learn from the sampled data. These cycles are repeated until a stopping criterion is reached, which can be a maximum number of cycles or a specific value in an evaluated metric (LORENA; GAMA; FACELI, 2011). In the testing stage, an unknown model is propagated through the network and computed by the layers until the output.

3.4.5 Random Forests

The RF method is a ML method that builds a set of DT, a directed acyclic graph where each node can be of division or leaf. The division nodes have a rule that evaluates data set

features which, depending on their result, forwards the flow to another division node or a leaf node. The leaf nodes define the class that the processed example belongs to (LORENA; GAMA; FACELI, 2011).

For improvement of the usual DT, it is common to use the combination of several individual classifiers, where the result of most DT gives the final result. This approach represents the function of the RF method, which combines several DT classifiers, whose final result consists of a majority vote (HO, 1995).

The RF algorithm introduces extra randomness when creating DT. The RF uses a predefined number (m) of features in training set to perform the divisions of the nodes. When a tree node is evaluated for splitting, only the randomly chosen m features are considered for the division node (DIETTERICH, 2000). Searching for random features for each tree node characterizes the embedded attribute selection technique.

In the test step, a new example is evaluated by the trees in the forest. The class assigned to the new example is the one that got the most votes, *i.e.* the class that had the most assignments.

3.5 Classifier Evaluation

The classifiers generated by the methods presented in Section 3.4 were evaluated using three approaches: confusion matrix (CM), k-fold cross-validation, and statistical hypothesis tests.

3.5.1 Confusion Matrix

The Confusion Matrix (CM) is a commonly used approach to evaluate predictive models. In Table 2, a binary classification CM is represented. The rows display the model's rating, while the columns represent the true prediction. The elements of CM cells can be defined by (THARWAT, 2020):

- *True Positives (TP)*: number of examples of the positive class correctly classified.
- *True Negatives (TN)*: quantity of predictions of the negative class correctly classified.
- *False Positives (FP)*: number of predictions of the positive class incorrectly classified.
- *False Negatives (FN)*: quantity of examples of the positive class incorrectly classified.

		True classification	
	Class	Positive	Negative
Model	Positive	TP	FP
Forecast	Negative	FN	TN

Tabela 2 – Binary classification confusion matrix.

This concept allows it to expand from a binary CM to a multiclass CM. In the multiclass evaluation of a CM, the values of TP, TN, FP, and FN vary according to the observed class (GRANDINI; BAGLI; VISANI, 2020). Table 3 represents a five class CM for evaluating class B individually. The intersection of column and row B represents the value of TP. The sum of row B, except for the intersection with column B (value of TP), forms the value of FP. The sum of column B, except the intersection of row B (value of TP), results in the value of FP. The sum of the other elements is the value of TN.

		True classification				
	Classes	A	B	C	D	E
Model Forecast	A	TN	FN	TN	TN	TN
	B	FP	TP	FP	FP	FP
	C	TN	FN	TN	TN	TN
	D	TN	FN	TN	TN	TN
	E	TN	FN	TN	TN	TN

Tabela 3 – Multiclass classification confusion matrix analyzing the values of TP, TN, FP and FN of class B.

From the CM, metrics can be extracted to evaluate the predictive models (HOSSIN; SULAIMAN, 2015), such as Accuracy (Acc.), Sensitivity (Sen.), Specificity (Spe.), Precision (Pre.), among others. The Acc., defined by Equation 27, represents the rate of samples correctly classified by the classifier, where i and j are the rows and columns of CM, respectively, and K is the number of problem classes.

$$Acc. = \frac{\sum_{i=1}^{K} CM_{i,i}}{\sum_{i=1}^{K} \sum_{j=1}^{K} CM_{i,j}}. \tag{27}$$

Using Equation (28), it is also possible to evaluate the hit rate per class (HRC) for a multiclass CM (OLIVA; ROSA, 2021). For two-class (binary) problems, HRC equals Acc..

$$HRC = \frac{TP + TN}{TP + FP + TN + FN}. \tag{28}$$

The concept of TP, TN, FP, and FN can be used to calculate other metrics commonly used in binary problems, such as:

- *Specificity (Spe.)*: percentage of negative samples correctly predicted as negative by the classifier is computed by Equation (29).

$$Spe. = \frac{TN}{TN + FP}. \tag{29}$$

- *Precision (Pre.)*: rate of predictions as a positive class was indeed positive and is defined by Equation (30).

$$Pre. = \frac{TP}{TP + FP}. \tag{30}$$

- *Sensitivity (Sen.)*: rate of positive samples correctly predicted as positive by the classifier. This measure can be calculated using Equation (31).

$$Sen. = \frac{TP}{TP + FN}. \tag{31}$$

Another metric commonly used in evaluating predictive models is the Macro F1-Score (MaF1), which evaluates the model considering all classes, as is done with the Acc. measure. This metric is interpreted as a harmonic mean of Sen. and Pre. (OPITZ; BURST, 2019). Firstly, the arithmetic mean of Pre. and Sen. is calculated using Equation (32) and (33), respectively. Finally, MaF1 can be defined by Equation (34).

$$\text{Mean Pre.} = \frac{\sum_{k=1}^{K} Pre._k}{K}. \tag{32}$$

$$\text{Mean Sen.} = \frac{\sum_{k=1}^{K} Sen._k}{K}. \tag{33}$$

$$MaF1 = 2 * \frac{\text{Mean Pre.} * \text{Mean Sen.}}{\text{Mean Pre.} + \text{Mean Sen.}}. \tag{34}$$

3.5.2 K-Fold Cross-Validation

Cross-validation is a technique that evaluates the generalization of a model from a data set associated with a ML method (KOHAVI *et al.*, 1995). This method helps in the evaluation of the model, in addition, to revealing the overfitting. The k-fold involves randomly dividing the data set into k groups, also called folds. One of the folds is separated for testing, and the others are used as training data. The cross-validation process is repeated k times so that each of the k groups is used once as a test data model (ANGUITA *et al.*, 2012).

This process will result in k evaluative metrics, one for each fold. The process obtains the metric to account for that model mean of the k-fold (REFAEILZADEH; TANG; LIU, 2009). Commonly, for ML problems, 10-fold ($k = 10$) are used to validate results because the average error values are close to those obtained when k is equal to the length of the data set (MARKATOU *et al.*, 2005).

3.5.3 Statistical Hypothesis Tests

Even with cross-validation, the survey of evaluative metrics is error-prone arising from chance and randomness (DEMŠAR, 2006). In this sense, statistical hypothesis tests are performed to verify differences and validate the results. Although these tests are not often applied to the classification of signs of epilepsy (OLIVA; ROSA, 2021), they can be essential for comparing

models. The main function of a statistical test for evaluating ML models is to verify whether the difference in ability to classify between two models is real or due to a statistical chance (DIETTERICH, 1998).

Different tests can be applied, but the choice of the appropriate method depends on several factors, such as presented next (DEMŠAR, 2006):

- *Distribution:* if the k cross-validation results are within the normal distribution;
- *Pairing:* also known as a dependency, indicates whether the models are trained from the same dataset.

The Shapiro-Wilk normality test (SHAPIRO; WILK, 1965) can be used to verify the distribution, whose main advantages include its easy application and recommendation for small samples (MIOT, 2017).

The assumption of a statistical test is called the null hypothesis, and one can calculate statistical measures and interpret them to decide whether the null hypothesis is accepted or rejected. In the case of the Shapiro-Wilk test, the null hypothesis is that the k-fold are normally distributed (SHAPIRO; WILK, 1965). If the test comes positive, the null hypothesis is accepted.

Statistical tests are used in this work to compare models with different input features (due to the feature selection). In this way, the data from our experiments will not have a dependency. In this case, the t-Student test is appropriate if the model results are within the normal distribution. Otherwise, the test that should be applied is the Mann-Whitney (KIM, 2015) test.

If the test result suggests insufficient evidence to reject the null hypothesis, then any observed difference in the model's skill is likely due to statistical chance. In other words, statistically, the two models have the same classification capacity. If the test result suggests that there is sufficient evidence to reject the null hypothesis, then the observed difference in the skill of the models is real (DEMŠAR, 2006).

Statistical hypothesis tests result in a $p - value$. It is said that there is a statistical significance or that the result is statistically significant when the observed $p - value$ is less than the significance α defined for the study (ZILIAK, 2017). For this type of study, the value of α is commonly 5%, *i.e.* the null hypothesis is accepted if the $p - value$ resulting from the test is less than 0.05.

4 EXPERIMENTAL EVALUATION AND RESULTS

Our experimental evaluation is divided into six steps (Figure 1), as the literature suggests:

1. EEG database acquisition (Section 4.1).
2. EEG segment representations (Section 4.2).
3. Feature extraction (Section 4.3).
4. Feature selection 4.4).
5. Classifier building (Section 4.5).
6. Classifier evaluation (Section 4.6).

4.1 EEG Database Acquisition

For the evaluation of the proposed models, the base Bonn EEG time series database [1] (ANDRZEJAK *et al.*, 2001), available to the public free of charge, was used in this work.

As the database has the widest class variety related to epilepsy, it is used in several studies on the application of ML methods for epilepsy detection(MURSALIN *et al.*, 2017; MOLLA *et al.*, 2020; VARGAS; OLIVA; TEIXEIRA, 2021; SHOEIBI *et al.*, 2021). This database is divided into five subsets of EEG segments from healthy volunteers and patients at different stages of seizures:

A: healthy volunteers with eyes open;

B: healthy volunteers with eyes closed;

C: the hippocampal formation of the opposite hemisphere of the brain in which the ictal phase of epilepsy is occurring. These segments are very similar to interictal signs;

D: the interictal activity of the epileptogenic zone of patients with epilepsy;

E: seizure activity, selected from all recording places that exhibit ictal activity from patients with epilepsy.

Each subset contains 100 single-channel EEG segments, represented as TS with a duration of 23.6 seconds, a sample rate of 173.61 Hz, and a resolution of 12 bits. In Figure 5, five samples of the Bonn database are illustrated, one for each subset. Some differences can be observed, such as segment B (normal with eyes closed), which contains more oscillations per second compared to segment A (normal with eyes open). Also, the segment of class E (ictal) has greater amplitudes than the others.

[1] http://bitly.ws/ooaQ

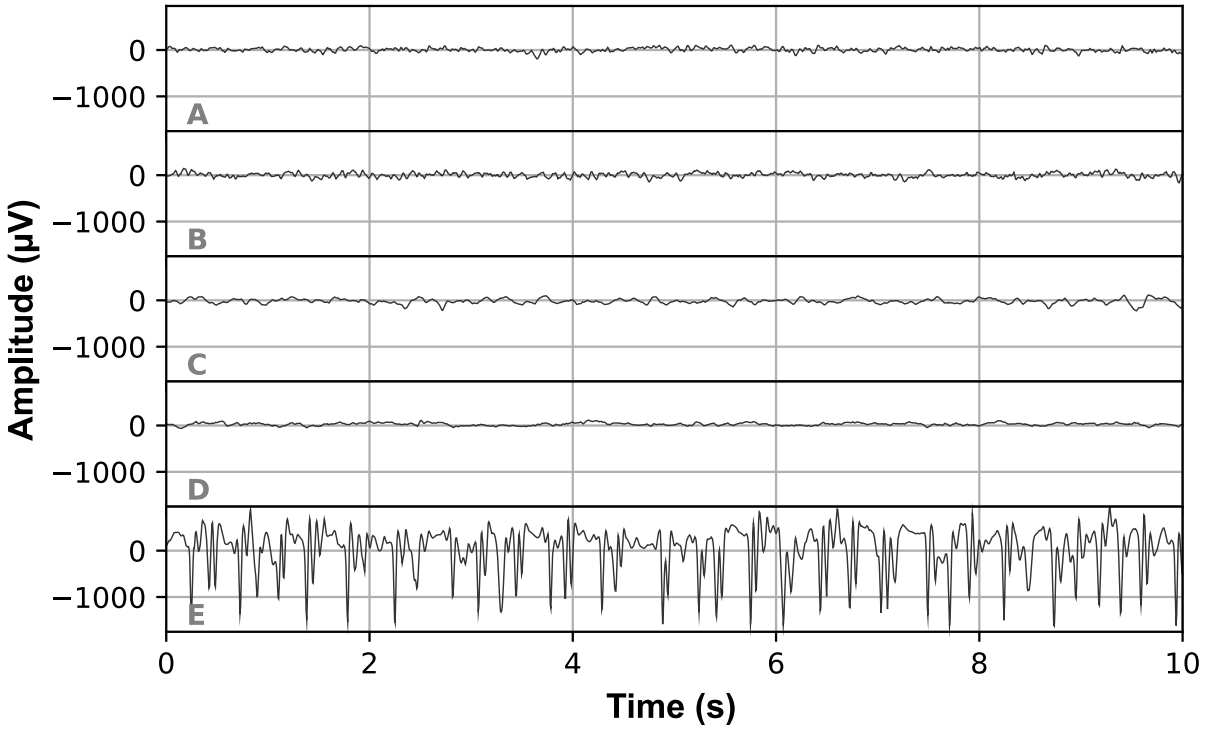


Figura 5 – Examples of EEG segments from the five subsets (A-E) of the Bonn database.

4.2 EEG Segment Representations

For each EEG segment, the multitaper method was applied to generate PS, SG, and BG, as described in Section 3.1.1, 3.1.2 and 3.1.3 respectively.

Figure 6 shows examples of PS from five samples, one from each set of the Bonn basis. The EEG E segment, ictal, contains the highest incidence of Alpha waves (8-12 Hz).

SG examples of five samples from the Bonn base are illustrated in Figure 7(a). In segments A, B, and D, a high magnitude band near 50Hz is observed, generated by interference from the electrical network while collecting the signals[2].

Five BG examples are illustrated in Figure 7(b). Comparing the representations, the E segment, ictal, presents higher magnitudes than the others.

Each Bonn database segment is represented as TS, PS, SG, and BG. These representations are used for feature extraction.

4.3 Feature Extraction

The third step corresponds to feature extraction from the EEG segments. After transforming the segments, each representation has specific features extracted, as shown in Section 3.2. The number of extracted features from each EEG segment is shown below:

[2] The example of EEG used to generate SG was collected in the city of Bonn, Germany, where the power grid is 50Hz

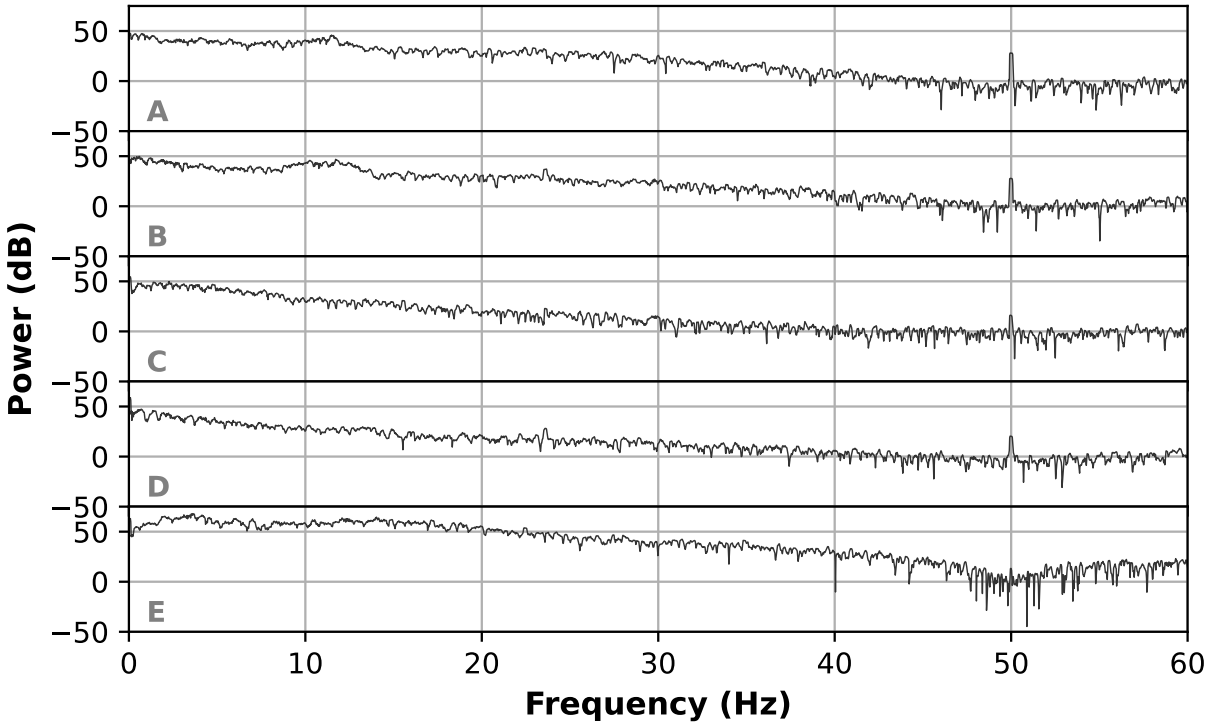
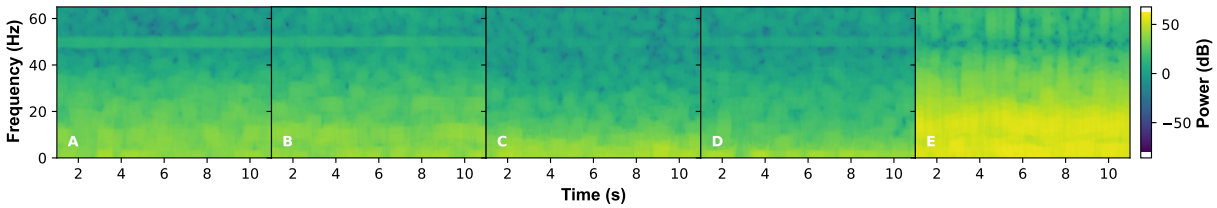
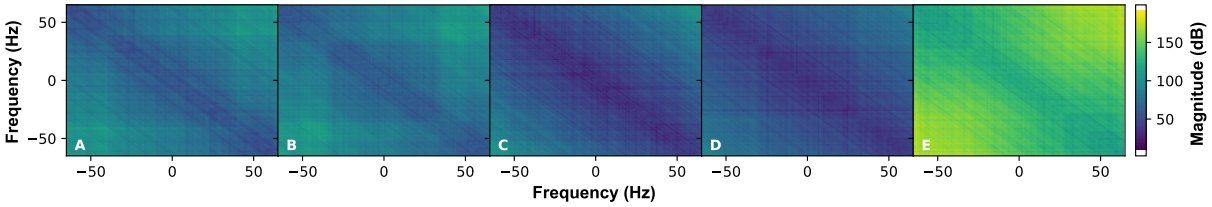


Figura 6 – PS examples of EEG segments from subsets of the Bonn database.



(a) Spectrogram



(b) Bispectrogram

Figura 7 – Examples of EEG representations of subsets from the Bonn database.

- *Time domain:* 25 features extracted from TS;
- *Frequency domain:* 156 features extracted from PS (26 from each sub-band and 26 from the entire spectrum);
- *Time-frequency domain:* 96 features extracted from SG (16 from each sub-band and 16 from the entire spectrum);
- *Nonlinear analysis:* 8 features of the complete BG.

Features extracted from each EEG representation are summarized in Table 4.

Tabela 4 – Features extracted from the EEG segment.

Time Serie	Power Spectrum	Spectrogram	Bispectrogram
Amplitude	Amplitude	Asymmetry	1st Order Diagonal Amplitude
Approximate Entropy	Approximate Entropy	Coefficient of variation	2nd Order Diagonal Amplitude
Asymmetry	Asymmetry	Crest Factor	Absolute Logarithmic Sum
Coefficient of variation	Coefficient of variation	Interquartile Size	Mean Magnitude
Crest factor	crest factor	Kurtosis	Normalized Entropy
Hjorth activity	Hurst exponent	Maximum	Normalized Quadratic Entropy
Hjorth complexity	Interquartile Size	Mean	X Weighted Center
Hjorth mobility	Kurtosis	Mean Square Value	Y Weighted center
Hurst exponent	Line Length	Minimum	
Interquartile Size	Maximum	Q1	
Kurtosis	Mean	Q2	
Line Length	Mean Square Value	Q3	
Maximum	Median	Spectral Band Energy	
Mean	Minimum	Spectral Band Width	
Mean Square Value	Mode	Spectral Centroid	
Median	Non-linear energy	Standard deviation	
Minimum	Peak Frequency		
Mode	Q1		
Non-Linear Energy	Q2		
Q1	Q3		
Q2	Renyi entropy		
Q3	Shannon entropy		
Renyi entropy	Spectral Centroid		
Shannon entropy	Spectral Entropy		
Standard deviation	Spectral Flatness		
	Standard deviation		

In total, 285 features were extracted from each EEG segment evaluated. However, this large quantity of measures can make the convergence of ML algorithms difficult. Compared to related works, a much smaller amount of features are used. In this case, feature selection methods are applied to estimate a suitable feature subset for building predictive models.

4.4 Feature Selection

This work proposes two feature selection techniques: one based on filter and another on the wrapper. They are detailed in this Section.

4.4.1 Filter by Pearson's Correlation Coefficient

The features were first analyzed by the filter method. The proposed algorithm was a CFS based on Pearson's correlation. A Pearson correlation coefficient matrix is generated to estimate the significance of each feature. Only features with more than 50% correlation with the output variable are selected for dimensionality reduction. This parameter setting is applied because it

was enough to represent the knowledge in the database, according to a previous study (still unpublished). In the next step, extracted features are compared, and those with a correlation greater than 50% are selected. Features with the lowest correlation with the output variable are excluded from the correlated pair.

Figure 8 illustrates results of the proposed filter method. This figure shows a reduction from 285 features to 37 in the first stage and 4 in the final stage.

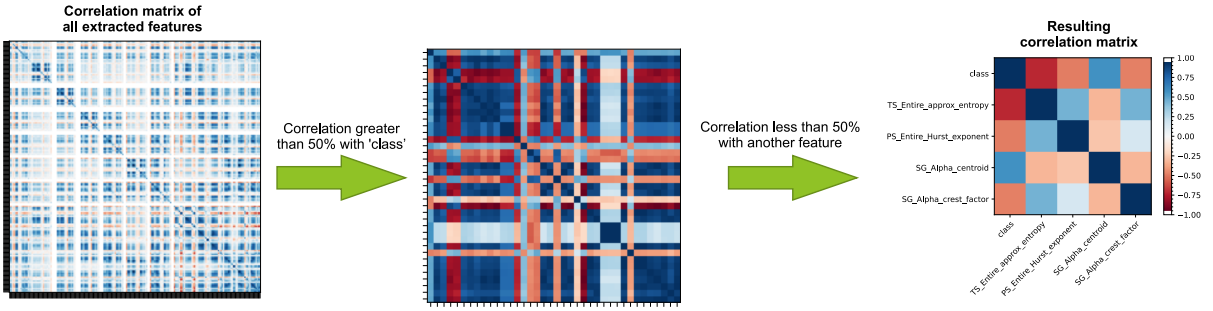


Figura 8 – Feature selection by Pearson's correlation coefficient.

These four features were evaluated by the methods of ML NB, KNN, SVM, and ANN. The settings for each algorithm are covered in Section 4.5. In the evaluation, ten-fold cross-validation is used. The NB model reached the lowest accuracy, 58.4%. Among the MLP classifiers, the model that stood out was the model with the RELU activation function in the intermediate layer, which obtained an accuracy of 67.8%. Of the KNN models, the one that achieved the highest classification was the one that used five nearest neighbours, which reached an accuracy of 73.6%. The SVM model, trained using a third-order polynomial kernel function, achieved the highest accuracy among all classifiers, which was 77.8%.

It is understood that the low amount of features, despite considerable correlation with the output variable, could not represent enough knowledge for ML algorithms to achieve satisfactory performance. Thus, increasing the feature amount, considering Pearson's correlation coefficient, could improve the performance of ML methods. For this, the minimum correlation threshold with the output variable was reduced from 50% to 40%, and measures were removed until a minimum quantity in the evaluated set was reached. This experiment was repeated for sets of 50, 40, 30, 20 and 10 features.

4.4.2 Wrapper Technique Based on Genetic Algorithm

Based on the technique presented in Section 3.3.2, an initial population of 90 individuals was used in the experimental evaluation. For classifier building, the third-order polynomial SVM is used since it was the setting that obtained the highest accuracy in the previous experiment. As a metric for evaluating the models, accuracy is used. The considered chance of an individual mutating is 5%. The difference between the best and the worst individual less than 0.01% is set

as a stopping criterion, so the chances of getting better samples than the current ones are almost non-existent. These parameters were determined by previous experiments (still not published).

Unlike what were done in the filter technique, cross-validation was not used in the wrapper. In this case, the data was divided into two sets, 80% for training and 20% for testing. As this method is tested several times, for each feature subset generation, a new evaluation is made by training an SVM model, resulting in different validation and test subsets. This approach is similar to the division that occurs in cross-validation, allowing fast convergence.

The wrapper technique was used to generate subsets of 4, 10, 20, 30, 40, and 50 features and compare them with those generated by the filter method. Figure 9 shows the difference, in accuracy, between the best and worst individuals from each subset, according to the generation evolution. The dashed black line represents the stopping criterion.

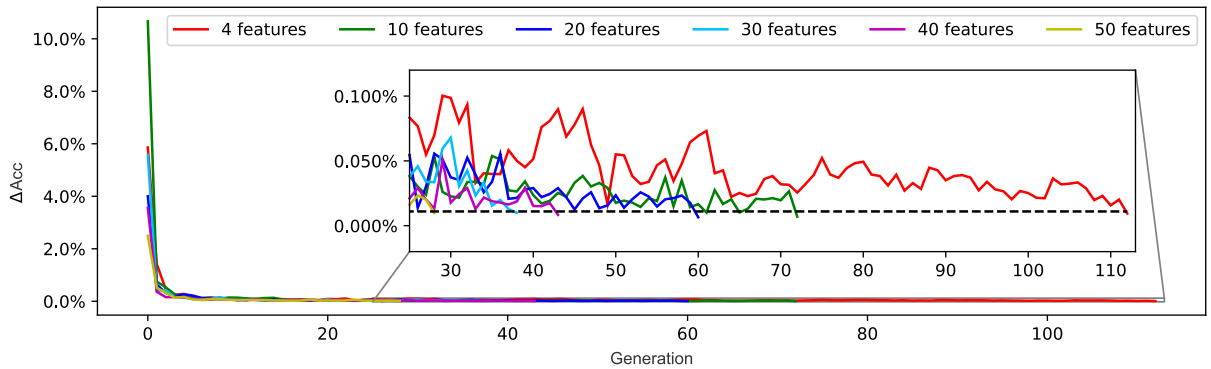


Figura 9 – Difference in accuracy between the best and worst individual of each population per generation.

The experiment with 50 features was the fastest to converge, with only 29 generations. The experiment with four features required more generations for convergence, which was 113. Apparently, the smaller the number of features, the more generations it takes for this algorithm to converge. However, the experiment with 30 features took fewer generations to converge than with 40 features. This fact probably occurred due to the algorithm having randomly made better choices at the time of reproduction or evolution of individuals.

4.4.3 Selected Features

The two feature selection methods used in this work presented distinct feature sets. In the filter selection, features from a subset of size four are kept until the selection of the size fifty subset. This operation is due to how the filter technique is implemented, always looking for the features that most correlate with the output class.

In the wrapper method, due to its randomness, the sets were formed by features that were not repeated often, forming highly distinct subsets.

Among the 285 features, 157 features were selected at least once by the proposed selection of feature methods. The comparison among the features selected by each method can be seen in the supplementary material[3].

4.5 Classifier building

The classifiers were constructed using the selected feature subsets as input to ML algorithms. In this work, variations of the methods described in Section 3.4 were discussed.

The model based on NB was the Multinomial Variation Naïve Bayes (RENNIE *et al.*, 2003), which adapts the versatility of the traditional algorithm for the classification of multiclass models. For the models based on KNN, described in the algorithm of Section 3.4.2, different ks were used: 1, 3, 5, and 7.

For SVM model building, the C-Support Vector Classification variation of libsvm (CHANG; LIN, 2011) is used. This tool is recommended for small samples (less than tens of thousands) and handles multiclass problems as one against one, decreasing the chances of the algorithm being spatialized in just one class. In the SVM approach, linear, polynomial (third, fifth, and seventh order polynomials), radial base function (RBF), and sigmoid function were evaluated because they are the most common in problems of this nature.

As a basis for the ANN models based on the MLP architecture, fully connected three-layer topologies were used. The input layer has the number of source nodes equal to the feature subset length. The output layer comprises five neurons with activation softmax, which returns a probability vector of the EEG segment of the network input referring to one of the five classes considered in this study. Only one middle (hidden) layer can implement any continuous function (CYBENKO, 1989) in addition to Heaton (2008)'s suggestion, in which this layer must have the number of two-thirds of the nodes input layer sources plus the quantity of output layer neurons.

In this article, six different activation functions for the middle (hidden) layer are discussed: hyperbolic tangent, sigmoid, softplus, Exponential Linear Unit (ELU), Scaled Exponential Linear Unit (SELU) and Rectified Linear Unit (RELU), as they were functions used in previous works that showed satisfactory results (VARGAS; OLIVA; TEIXEIRA, 2021).

As the algorithm's stopping criterion metric, the mean square error (MSE) was used, defined according to Equation (35), where n is the number of samples, Y is the true value of the output, and $\bar{Y}$ is the predicted value. The higher the MSE, the worse the model performance. As a stopping criterion, during susceptible iterations of the training algorithm, MSE could not have a reduction in the subsequent ten iterations.

$$MSE = \frac{1}{n}\sum_{i-1}^{n}(Y_i - \bar{Y}_i)^2. \tag{35}$$

[3] https://ddsl.me/vZy7AYr

Although some ML algorithms can use embedded methods, such as ANN that implement L1 regularization to select features, these parameters were not changed in the implementation. So, only RF models are considered embedded in our experiments.

The 17 model variations (one NB, four KNN, six SVM, and six MLP) are tested for the 12 feature subsets selected by the selection techniques (six by the filter and six by the wrapper) proposed. So, 204 models were built for feature selection evaluation. On the other hand, 60 RF models were built, whose variations change the depth of the tree and the number of features used per node. RF techniques internally use built-in feature selection. Models that use the entire set of features, that is, do not use dimensionality reduction techniques, were also analyzed to compare the results.

In total, 281 variations of classification models were evaluated[4].

4.6 Classifier Evaluation

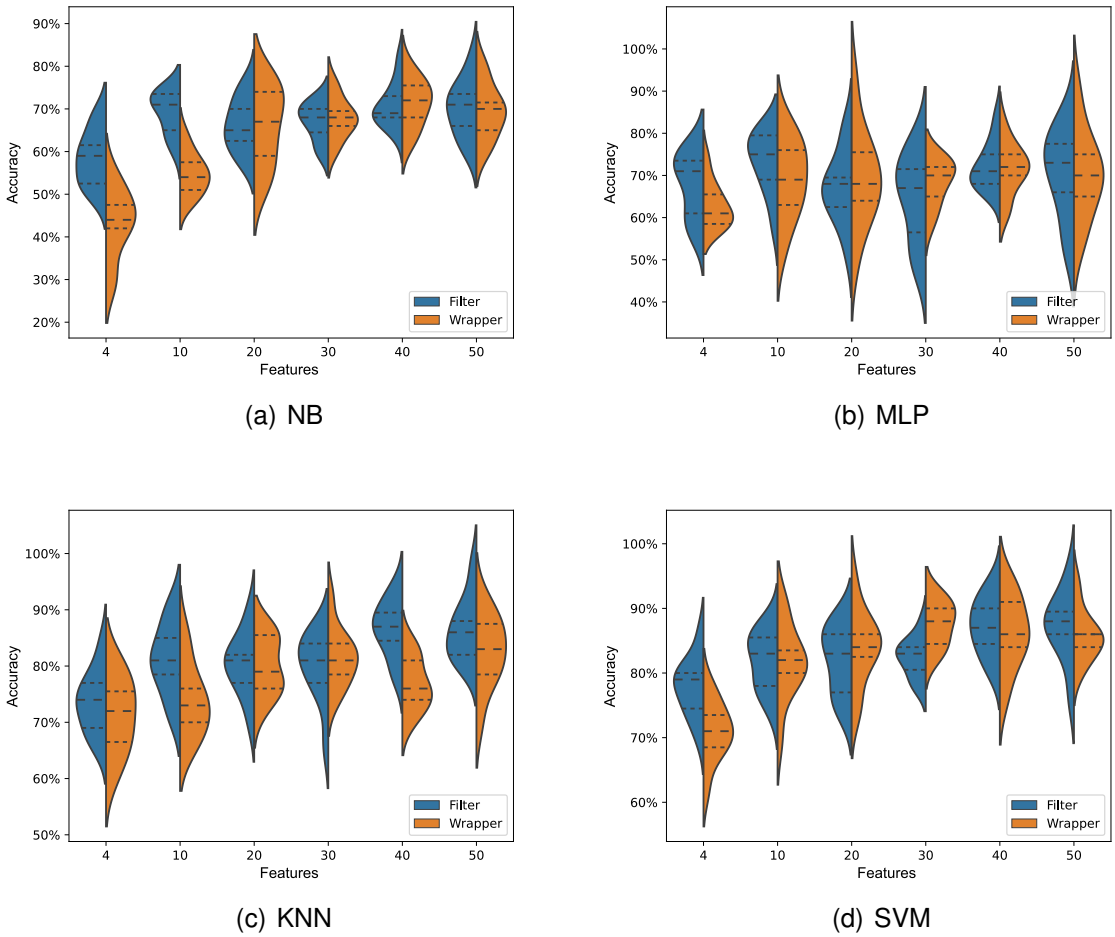
ML classifiers built from the selected feature subsets are evaluated using the ten-fold cross-validation technique. Figure 10 presents the accuracy distribution (Y-axis) of the most accurate model for each ML algorithm, considering the length of each feature subset (X-axis). In this figure, the long dashed line represents the second quartile, and the short dashed lines represent the first and third quartiles of the ten-fold.

For NB models (Figure 10(a)), the accuracy increases when the amount of selected features increases for both attribute selection techniques. In most cases, the filter-based method's performance reached superior accuracy compared to the wrapper. The filter-based model built from a subset with four features achieved the worst mean accuracy, which was 58.4%. The highest mean accuracy filter-based model (70.8%) is built from a subset with 40 features. However, the models that used features selected by the wrapper method achieved the highest (71.2%) and lowest (43.2%) accuracy for the NB algorithm when applied in 40 and four features, respectively.

The classifiers based on MLP (Figure 10(b)) obtained the lowest variation of mean accuracy (approximately 10%) among the ML algorithms used in this work. The filter-based models obtained a slightly superior performance compared to the wrapper-based models. The filter-based model built from ten selected features and set with ELU activation function achieved the highest mean accuracy (73.2%). The most accurate wrapper-based model (mean accuracy of 71.4%) was constructed using 40 selected features and set with SELU activation function.

For the KNN algorithm (Figure 10(c)), the filter-based model achieved the highest average accuracy (86.6%). It was built with 40 features, considering five nearest neighbours during classification. The wrapper-based model with the highest average accuracy (82.6%) was trained using 50 features, considering seven nearest neighbours.

[4] 102 models with feature selection by the filter method, 102 models with feature selection by the wrapper method, 60 models based on the RF algorithm, and 17 models without feature selection



(a) NB (b) MLP

(c) KNN (d) SVM

Figura 10 – 10-fold distribution for the proposed filter and wrapper feature selection techniques.

Among the algorithms presented so far, the models that used SVM were the ones that obtained the best results, mainly using the third-order polynomial kernel. These models reached the better metrics, except one built from a subset with four features. The wrapper-based models obtained mean accuracy values between 71% (four feature subset) and 87.6% (30 feature subset). The filter-based classifiers reached mean accuracy of 87.2% (the highest in this setting) for subsets with 40 and 50 features. As a tie-breaking criterion, the filter-based model with 40 features performs better because it needs fewer computational resources to achieve the same result compared to one built from 50 features.

The results presented in Figure 10 suggest that, among all the models, the ML algorithm that obtained the best metrics was the SVM with third order polynomial kernel for both selection techniques of features. For the filter method, a mean accuracy of 87.2% is reached. For the wrapper method, a mean accuracy of 87.6% is achieved.

It is also noteworthy that in most of the evaluated algorithms, the distribution and the interquartile distance of the models with 30 and 40 features are more compact. This distribution directly impacts the accuracy standard deviation. Thus, the results are less distributed in the

evaluated folds. In this sense, our experimental evaluation suggests that more homogeneous and representative features belong to the sets of 30 and 40 features.

RF-based models were also evaluated. The RF models were taken into account to compare with the highest results of the Filter and Wrapper methods. Due to how RF selects the characteristics, the models generated based on RF did not obtain great variation between their results (less than 1% between the highest and lowest average accuracy of the 10 folds), obtaining their best performance with a forest of 40 trees and an average accuracy of 90.99%.

Additionally, models generated using attribute extraction methods are compared to classifiers built using all 285 features. This comparison aims to verify whether the attribute selection method improved the performance of ML algorithms in our study. The accuracy distribution of the models that obtained the best metrics is presented in Figure 11.

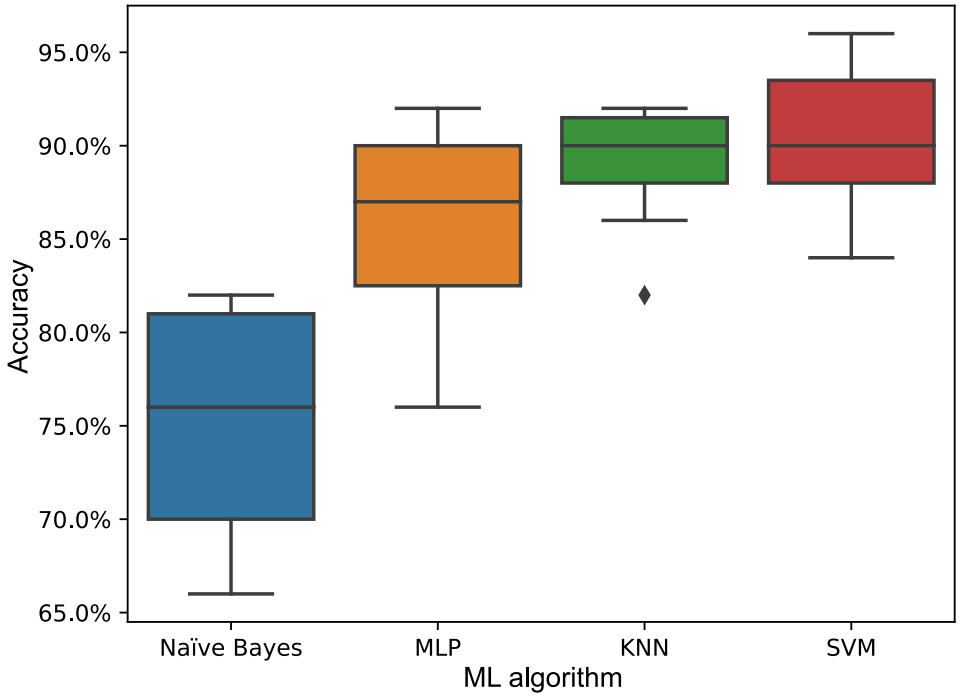


Figura 11 – 10-fold distribution for all extracted features compared by the ML methods used in this study.

According to Figure 11, the NB algorithm obtained the lowest mean accuracy, reaching 74.8% accuracy. The MLP classifier that obtained the highest accuracy (82.2%) was the one that used the ELU activation function. The KNN model, with a 1-nearest neighbour, reached a mean accuracy of 89.0% and the lowest standard deviation (3.16%). The best metrics were obtained again by the SVM models trained using a seventh-order polynomial kernel, reaching a mean accuracy of 90.2%.

The evaluative metrics by class, according to Section 3.5.1, of the models that obtained the highest average accuracy by attribute selection technique are presented in Table 5.

		A	B	C	D	E
HRC	Filter	96.40	96.80	91.60	90.80	98.80
	Wraper	97.40	97.60	91.60	90.60	98.00
	Embeding	97.40	97.20	94.60	94.00	98.80
	Without selection	98.20	98.40	92.60	92.20	99.00
Spe.	Filter	97.18	98.26	92.50	96.18	99.76
	Wraper	97.74	98.48	93.23	95.75	99.23
	Embeding	97.49	98.75	95.57	97.55	99.31
	Without selection	98.06	99.74	95.29	95.18	99.52
Sen.	Filter	93.02	91.14	87.74	67.69	94.27
	Wraper	95.17	94.43	83.71	70.90	93.38
	Embeding	97.00	90.51	90.64	80.86	97.56
	Without selection	98.89	94.01	80.77	79.39	97.11
Pre.	Filter	90.74	93.45	74.56	80.63	99.00
	Wraper	91.31	94.12	75.14	81.12	97.06
	Embeding	90.99	94.21	85.51	88.15	95.83
	Without selection	92.44	99.17	80.68	79.69	97.75

Tabela 5 – Percentage(%) of HRC, Pre., Sen., and Spe. by class of the best ML models for attribute selection techniques and without selection.

In this table, it is possible to visualize the relationship between classifiers and classes. All four models present larger HRC for class E. These results demonstrate that these models can more easily distinguish the samples from the ictal segments, which refer to the moment of the crisis, as they present higher metrics in the E class. The lowest values are observed for Sen. and Pre. of class D models. These values occur because several segments of classes C and D have high similarities, making their differentiation difficult.

With the Table 5 it is possible to compute the metrics as presented in Section 3.5.1 which are represented by Table 6. The RF model stands out in all the metrics evaluated, even surpassing the accuracy obtained by the model without feature selection. It is worth noting that the filter-based and wrapper-based models obtained similar results, although the model that uses the wrapper filter reached slightly higher metrics. Also, there is a variation of less than 5% between the metrics of the four models. This fact highlights the possibility that the models have the same selection capability. Additionally, the RF model results must be highlighted since they surpassed those obtained by the model without attribute selection.

Features Selection	Selected Features	ML Algorithm	Model Accuracy	Average by Class				Macro F1-Score
				HRC	Spe.	Sen.	Pre.	
Filter	40	SVM	87.20%	94.88%	96.78%	86.77%	87.67%	87.22%
Wraper	30	SVM	87.60%	95.04%	96.89%	87.52%	87.75%	87.63%
Embeding	-	RF	90.99%	96.40%	97.74%	91.32%	90.94%	91.13%
Without selection	285	SVM	90.20%	96.08%	97.56%	90.03%	89.95%	89.99%

Tabela 6 – Accuracy, macro F1-Score, averages of HRC, specificity, precision, and sensitivity of the models with the best performance.

Despite the difference in metrics between the models, it is not yet possible to prove that a model has a superior performance in relation to others. For this, hypothesis statistical tests

Feature	p-Value	
Selection Method	Shapiro-Wilk	t-Student
Filter	0.95347	0.14513
Wraper	0.47262	0.15914
Embeding	0.39603	0.70219
Without selection	0.41498	-

Tabela 7 – Results of the Shapiro-Wilk and t-Student tests.

must be conducted, whose chosen methods and their respective results are presented in the next section.

4.6.1 Statistical Evaluation

Two tests were applied to verify if there is a statistically significant difference between the models: the Shapiro-Wilk test; and the t-Student or Mann-Whitney test, depending on the result of the first test.

Firstly, it was verified if the distribution of the accuracy of the k-fold behaved like a normal distribution. Although Figure 10 and 11 represent, approximately, a normal distribution for the models that obtained the best result by the selection of feature technique, this conclusion can only be reached with a test of normality, like Shapiro-Wilk.

Table 7 presents the results of the Shapiro-Wilk test application. In all cases, the null hypothesis is rejected. Thus, the alternative hypothesis is accepted that accuracy values are normally distributed. As the samples are normally distributed, the t-Student test is chosen. In this case, the null hypothesis suggests that differences between the models are merely due to randomness. In contrast, the alternative hypothesis suggests that the difference among models is due to their capacity to classify. After the statistical hypothesis testing, it was stated that the null hypothesis was accepted. Therefore, the models have the same classification capacity.

The t-Student test was also used to compare the two proposed selection of feature methods, such as Pearson correlation (filter) and genetic algorithm (wrapper). As a result, this test reached a p-value of 0.82558. In this way, the null hypothesis that the two models may have different results is also rejected, with 5% significance.

4.6.2 Comparison of Results

In this Section, our results are compared with related studies. All compared works presented the classification of the five classes, *i.e.* the same database is used. Table 8 compares the following metrics: sensitivity, specificity, macro F1-Score, and accuracy.

Our result surpassed the metrics of related works, exposing a classification potential similar to those considered state of the art. Most existing works surpass the classification capacity of the models proposed in this work using deep learning techniques. These approaches have

	Spe.	Sen.	MaF1	Acc.
Tzallas, Tsipouras e Fotiadis (2009)	89.10%	89.00%	-	89.00%
Liang, Wang e Chang (2010)	-	-	-	85.90%
Nicolaou e Georgiou (2012)	93.23%	94.38%	-	86.10%
Ramakrishnan e Murugavel (2019)	93.00%	97.00%	-	95.00%
Oliva e Rosa (2019)	-	-	-	86.00%
Tsipouras (2019)	-	-	-	91.20%
Türk e Özerdem (2019)	98.33%	93.60%	93.57%	93.60%
Li *et al.* (2020)	-	-	94.59%	94.60%
Vargas, Oliva e Teixeira (2021)	96.05%	84.45%	85.20%	84.20%
Filter	96.78%	86.77%	87.22%	87.20%
Wraper	96.89%	87.52%	87.63%	87.60%
Embeding	97.74%	91.32%	91.13%	90.99%
Without selection	97.56%	90.03%	89.99%	90.20%

Tabela 8 – Comparison of the results obtained with works found in the literature for the classification of 5 classes of the Bonn database.

difficult explainability and reproducibility and require large databases for training (KOHLI *et al.*, 2016; MARCUS, 2020). Thus, it is argued that although they show compelling results, the practical applicability of these models may expose a significant gap.

It is also possible to point out that some related works do not use cross-validation methods such as Nicolaou e Georgiou (2012), which only divides the samples into 60% for training and 40% for testing. This approach is similar to using only one fold. In comparison, the analysis of the results separately from the folds in the experiments previously conducted by us (still not published) showed accuracy of 96%, surpassing that obtained in all the related works mentioned.

Other works such as Tzallas, Tsipouras e Fotiadis (2009) use unconventional cross-validation. A division of 10 folds is done, but the division between training and testing is 50% randomly. This method can eventually use similar training sets by adjusting the models to the exposed data.

Other references, like Ramakrishnan e Murugavel (2019) and Türk e Özerdem (2019), make use of the ten-fold cross-validation, but it is not clear how the training and testing sets are divided. In this way, the direct comparison of the results is inconclusive since the validation methods were conducted differently and directly affected the final result.

It has not yet been possible to make a comparison through statistical tests with related works because, in general, only the mean of cross-validation folds are exposed, compromising the comparative stage.

5 CONCLUSION

In this article, predictive models were built to classify EEG segments among five different classes. For this, firstly, PS, SG, and BG are generated by the multitaper method, from each EEG segment. A total of 285 measurements were extracted from the EEG representations. Subsequently, features were selected by two proposed techniques, and the resulting subsets were evaluated by classifiers constructed from variations of ML algorithms. The evaluation of the results was carried out through cross-validation, confusion matrix, and statistical hypothesis tests.

Although it was impossible to define the best classifier for epilepsy detection, the experimental evaluations reached competitive results concerning previously published works, which are achieved mainly by models that use deep learning approaches. Even though many methods based on deep learning have been used in recent research, conventional ML methods are still effective for classification problems, as they demand smaller databases and less computational power. Furthermore, training efficient classifiers require only a small and representative database instead of extensive and commonly redundant ones. It is also believed that the performance of the proposed predictive models can be improved by extracting new features, using other ML algorithms to construct classifiers, and adjusting parameters of already used ML algorithms.

The models proposed in this work benefit from using feature selection by computational techniques, thus facilitating signal processing by ML methods. The experimental results confirm that the proposed methods reduce the computational effort by considerably reducing the number of features extracted from the EEG segments to obtain equivalently equal results.

5.1 Contributions

EEG segments in the form of time series were converted into three representations used in the literature through variations of the multitaper method. A total of 285 features were extracted from the EEG.

This work used feature selection techniques to assemble representative subsets that 281 predictive models evaluated.

With a reduced but representative set of characteristics, it was possible to achieve significant results, reaching metrics close to works considered state of the art.

5.2 Limitations

The database used is limited to only one channel. Encephalographic equipment acquisitions make use of multiple data channels in the exam process. It was not possible to statistically

compare related works because, in a usual way, the works do not present the results in the form of folds. Only the average of the folds is presented as the final result.

5.3 Future works

Future works include the implementation of other measures for feature extraction, new feature selection techniques, and using different ML algorithms to build classifiers. The evaluated models can also be used in more complex databases, such as epileptic signal databases with more channels.

REFERÊNCIAS

ACHARYA, U. R.; SREE, S. V.; SURI, J. S. Automatic detection of epileptic eeg signals using higher order cumulant features. **International journal of neural systems**, World Scientific, v. 21, n. 05, p. 403–414, 2011.

ACHARYA, U. R. *et al.* Automated eeg analysis of epilepsy: A review. **Knowledge-Based Systems**, v. 45, p. 147–165, 2013. ISSN 0950-7051.

AHA, D. W.; KIBLER, D.; ALBERT, M. K. Instance-based learning algorithms. **Machine learning**, Springer, v. 6, n. 1, p. 37–66, 1991.

ANDRZEJAK, R. G. *et al.* Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. **Physical Review E**, APS, v. 64, n. 6, p. 061907, 2001.

ANGUITA, D. *et al.* The 'k'in k-fold cross validation. *In*: I6DOC. COM PUBL. **20th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)**. [*S.l.*], 2012. p. 441–446.

BABADI, B.; BROWN, E. N. A review of multitaper spectral analysis. **IEEE Transactions on Biomedical Engineering**, IEEE, v. 61, n. 5, p. 1555–1564, 2014.

BABATUNDE, O. *et al.* A genetic algorithm-based feature selection. **International Journal of Electronics Communication and Computer Engineering**, v. 5, p. 889–905, 07 2014.

BELLMAN, R. Dynamic programming. **Science**, v. 153, n. 3731, p. 34–37, 1966.

BLUM, A. L.; LANGLEY, P. Selection of relevant features and examples in machine learning. **Artificial Intelligence**, v. 97, n. 1, p. 245–271, 1997. ISSN 0004-3702. Relevance.

BRIGHAM, E. O. **The fast Fourier transform and its applications**. [*S.l.*]: Prentice-Hall, Inc., 1988.

CHANG, C.-C.; LIN, C.-J. Libsvm: a library for support vector machines. **ACM transactions on intelligent systems and technology (TIST)**, Acm New York, NY, USA, v. 2, n. 3, p. 1–27, 2011.

COHEN, L. **Time-frequency analysis**. [*S.l.*]: Prentice hall New Jersey, 1995. v. 778.

CORTES, C.; VAPNIK, V. Support-vector networks. **Machine learning**, Springer, v. 20, n. 3, p. 273–297, 1995.

CURA, O. K.; AKAN, A. Analysis of epileptic EEG signals by using dynamic mode decomposition and spectrum. **Biocybernetics and Biomedical Engineering**, Elsevier, v. 41, n. 1, p. 28–44, 2021.

CYBENKO, G. Approximation by superpositions of a sigmoidal function. **Mathematics of control, signals and systems**, Springer, v. 2, n. 4, p. 303–314, 1989.

DEMŠAR, J. Statistical comparisons of classifiers over multiple data sets. **The Journal of Machine Learning Research**, JMLR. org, v. 7, p. 1–30, 2006.

DIETTERICH, T. G. Approximate statistical tests for comparing supervised classification learning algorithms. **Neural computation**, MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info . . . , v. 10, n. 7, p. 1895–1923, 1998.

DIETTERICH, T. G. Ensemble methods in machine learning. *In*: SPRINGER. **International workshop on multiple classifier systems**. [*S.l.*], 2000. p. 1–15.

DUBNOV, S. Generalization of spectral flatness measure for non-gaussian linear processes. **IEEE Signal Processing Letters**, IEEE, v. 11, n. 8, p. 698–701, 2004.

FISHER, R. S.; SCHARFMAN, H. E.; DECURTIS, M. **How Can We Identify Ictal and Interictal Abnormal Activity?** Dordrecht: Springer Netherlands, 2014. 3–23 p. ISBN 978-94-017-8914-1.

FIX, E.; HODGES, J. L. Discriminatory analysis. nonparametric discrimination: Consistency properties. **International Statistical Review/Revue Internationale de Statistique**, JSTOR, v. 57, n. 3, p. 238–247, 1989.

FREEDMAN ROBERT PISANI, R. P. D. **Statistics**. 4th. ed. [*S.l.*]: W. W. Norton & Company, 2007. ISBN 0393929728,978-0393929720.

FREEMAN, W.; QUIROGA, R. Q. **Imaging brain function with EEG: advanced temporal and spatial analysis of electroencephalographic signals**. [*S.l.*]: Springer Science & Business Media, 2012.

FREITAS, A. A. **Data mining and knowledge discovery with evolutionary algorithms**. [*S.l.*]: Springer Science & Business Media, 2002.

GHAYAB, H. R. A. *et al.* Classification of epileptic eeg signals based on simple random sampling and sequential feature selection. **Brain informatics**, SpringerOpen, v. 3, n. 2, p. 85–91, 2016.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep learning**. [*S.l.*]: MIT press, 2016.

GRANDINI, M.; BAGLI, E.; VISANI, G. Metrics for multi-class classification: an overview. **arXiv preprint arXiv:2008.05756**, 2020.

GUYON, I.; ELISSEEFF, A. An introduction to feature extraction. *In*: **Feature extraction**. [*S.l.*]: Springer, 2006. p. 1–25.

HALL, M. A. Correlation-based feature selection of discrete and numeric class machine learning. **Computer Science Working Papers**, University of Waikato, Department of Computer Science, 2000.

HASSAN, K. M. *et al.* Epileptic seizure detection from eeg signals using multiband features with feedforward neural network. *In*: IEEE. **2019 International Conference on Cyberworlds (CW)**. [*S.l.*], 2019. p. 231–238.

HASTIE, T. *et al.* **The elements of statistical learning: data mining, inference, and prediction**. [*S.l.*]: Springer, 2009. v. 2.

HAYKIN, S. **Neural Networks and Learning Machines**. 3rd. ed. [*S.l.*]: Prentice Hall, 2008. ISBN 0131471392.

HEARST, M. A. Trends & controversies: Support vector machines. **IEEE Intell. Syst.**, v. 13, p. 18–28, 1998.

HEATON, J. **Introduction to neural networks with Java**. [*S.l.*]: Heaton Research, Inc., 2008.

HEATON, J. An empirical analysis of feature engineering for predictive modeling. *In*: **SoutheastCon 2016**. [*S.l.*: *s.n.*], 2016. p. 1–6.

HERNÁNDEZ, D. *et al.* Detecting epilepsy in eeg signals using time, frequency and time-frequency domain features. *In*: **Computer science and engineering—theory and applications**. [*S.l.*]: Springer, 2018. p. 167–182.

HJORTH, B. Eeg analysis based on time domain properties. **Electroencephalography and clinical neurophysiology**, Elsevier, v. 29, n. 3, p. 306–310, 1970.

HO, T. K. Random decision forests. *In*: IEEE. **Proceedings of 3rd international conference on document analysis and recognition**. [*S.l.*], 1995. v. 1, p. 278–282.

HOSSEINZADEH, D.; KRISHNAN, S. Combining vocal source and mfcc features for enhanced speaker recognition performance using gmms. *In*: **2007 IEEE 9th Workshop on Multimedia Signal Processing**. [*S.l.*: *s.n.*], 2007. p. 365–368.

HOSSIN, M.; SULAIMAN, M. N. A review on evaluation metrics for data classification evaluations. **International journal of data mining & knowledge management process**, Academy & Industry Research Collaboration Center (AIRCC), v. 5, n. 2, p. 1, 2015.

HWANG, S. T.; GOODMAN, T.; STEVENS, S. J. Painful seizures: a review of epileptic ictal pain. **Current pain and headache reports**, Springer, v. 23, n. 11, p. 1–7, 2019.

JOHN, G. H.; KOHAVI, R.; PFLEGER, K. Irrelevant features and the subset selection problem. *In*: **Machine learning proceedings 1994**. [*S.l.*]: Elsevier, 1994. p. 121–129.

KEETON, G. **What is crest factor and why is it important?** 2015. Disponível em: https://encr.pw/lDA1n. Acesso em: 09 abr. 2021.

KHALID, S.; KHALIL, T.; NASREEN, S. A survey of feature selection and feature extraction techniques in machine learning. *In*: **2014 Science and Information Conference**. [*S.l.*: *s.n.*], 2014. p. 372–378.

KIM, T. K. T test as a parametric statistic. **Korean journal of anesthesiology**, Korean Society of Anesthesiologists, v. 68, n. 6, p. 540, 2015.

KOHAVI, R.; JOHN, G. H. Wrappers for feature subset selection. **Artificial Intelligence**, v. 97, n. 1, p. 273–324, 1997. ISSN 0004-3702. Relevance.

KOHAVI, R. *et al.* A study of cross-validation and bootstrap for accuracy estimation and model selection. *In*: MONTREAL, CANADA. **Ijcai**. [*S.l.*], 1995. v. 14, p. 1137–1145.

KOHLI, N. *et al.* Hierarchical representation learning for kinship verification. **IEEE Transactions on Image Processing**, IEEE, v. 26, n. 1, p. 289–302, 2016.

KRAMER, C.; GERHARDT, H. J. **Advances in wind engineering**. [*S.l.*]: Elsevier, 2012.

KUHN, M.; JOHNSON, K. *et al.* **Applied predictive modeling**. [*S.l.*]: Springer, 2013. v. 26.

LAL, T. N. *et al.* Embedded methods. *In*: **Feature extraction**. [*S.l.*]: Springer, 2006. p. 137–165.

LI, D. *et al.* A sequential method using multiplicative extreme learning machine for epileptic seizure detection. **Neurocomputing**, v. 214, p. 692–707, 2016. ISSN 0925-2312.

LI, L. M. *et al.* Demonstration project on epilepsy in brazil: situation assessment. **Arquivos de neuro-psiquiatria**, SciELO Brasil, v. 65, p. 5–13, 2007.

LI, X. **Signal processing in neuroscience**. [*S.l.*]: Springer, 2016. 288 p. ISBN 9811018219,978-9811018213.

LI, Y. *et al.* Epileptic seizure detection in eeg signals using a unified temporal-spectral squeeze-and-excitation network. **IEEE Transactions on Neural Systems and Rehabilitation Engineering**, IEEE, v. 28, n. 4, p. 782–794, 2020.

LIANG, S.-F.; WANG, H.-C.; CHANG, W.-L. Combination of eeg complexity and spectral analysis for epilepsy diagnosis and seizure detection. **EURASIP journal on advances in signal processing**, Springe, v. 2010, p. 1–15, 2010.

LIU, Z.; WANG, Y.; CHEN, T. Audio feature extraction and analysis for scene segmentation and classification. **Journal of VLSI signal processing systems for signal, image and video technology**, Springer, v. 20, n. 1, p. 61–79, 1998.

LORENA, A. C.; GAMA, J.; FACELI, K. **Inteligência artificial: uma abordagem de aprendizado de máquina**. [*S.l.*]: Grupo Gen-LTC, 2011. ISBN 8521618808.

MANOLIS, T. A. *et al.* Sudden unexpected death in epilepsy: The neuro-cardio-respiratory connection. **Seizure**, v. 64, p. 65–73, 2019. ISSN 1059-1311.

MARCUS, G. The next decade in ai: four steps towards robust artificial intelligence. **arXiv preprint arXiv:2002.06177**, 2020.

MARKATOU, M. *et al.* Analysis of variance of cross-validation estimators of the generalization error. **Journal of Machine Learning Research 6**, p. 1127–1168, 2005.

MARTÍNEZ-VARGAS, J. D.; GODINO-LLORENTE, J. I.; CASTELLANOS-DOMINGUEZ, G. Time–frequency based feature selection for discrimination of non-stationary biosignals. **EURASIP Journal on Advances in Signal Processing**, Springer, v. 2012, n. 1, p. 1–18, 2012.

MEHLA, V. K. *et al.* An efficient method for identification of epileptic seizures from eeg signals using fourier analysis. **Physical and Engineering Sciences in Medicine**, Springer, v. 44, n. 2, p. 443–456, 2021.

MIOT, H. A. **Avaliação da normalidade dos dados em estudos clínicos e experimentais**. SciELO Brasil, 2017. 88–91 p. Disponível em: https://www.t.ly/D4ip.

MITCHELL, T. M. *et al.* **Machine learning**. [*S.l.*]: McGraw-hill New York, 1997.

MOLLA, M. *et al.* Graph eigen decomposition-based feature-selection method for epileptic seizure detection using electroencephalography. **Sensors**, Multidisciplinary Digital Publishing Institute, v. 20, n. 16, p. 4639, 2020.

MURSALIN, M. *et al.* Automated epileptic seizure detection using improved correlation-based feature selection with random forest classifier. **Neurocomputing**, Elsevier, v. 241, p. 204–214, 2017.

NICOLAOU, N.; GEORGIOU, J. Detection of epileptic electroencephalogram based on permutation entropy and support vector machines. **Expert Systems with Applications**, Elsevier, v. 39, n. 1, p. 202–209, 2012.

NIKIAS, C.; MENDEL, J. Signal processing with higher-order spectra. **IEEE Signal Processing Magazine**, v. 10, n. 3, p. 10–37, 1993.

OLIVA, J. T. **Geração automática de laudos médicos para o diagnóstico de epilepsia por meio do processamento de eletroencefalogramas utilizando aprendizado de máquina**. 2019. Tese (Doutorado) — Universidade de São Paulo, 2019.

OLIVA, J. T. *et al.* Prototype system for feature extraction, classification and study of medical images. **Expert Systems with Applications**, Elsevier, v. 63, p. 267–283, 2016.

OLIVA, J. T.; ROSA, J. L. G. Predictive models for differentiation between normal and abnormal eeg through cross-correlation and machine learning techniques. *In*: **Towards Integrative Machine Learning and Knowledge Extraction**. [*S.l.*]: Springer, 2017. p. 134–145.

OLIVA, J. T.; ROSA, J. L. G. Classification for EEG report generation and epilepsy detection. **Neurocomputing**, Elsevier, v. 335, p. 81–95, 2019.

OLIVA, J. T.; ROSA, J. L. G. Binary and multiclass classifiers based on multitaper spectral features for epilepsy detection. **arXiv e-prints**, p. arXiv–2004, 2020.

OLIVA, J. T.; ROSA, J. L. G. Binary and multiclass classifiers based on multitaper spectral features for epilepsy detection. **Biomedical Signal Processing and Control**, v. 66, p. 102469, 2021. ISSN 1746-8094.

OPITZ, J.; BURST, S. Macro f1 and macro f1. **arXiv preprint arXiv:1911.03347**, 2019.

PAGE, A. *et al.* Comparing raw data and feature extraction for seizure detection with deep learning methods. *In*: **The Twenty-Seventh International Flairs Conference**. [*S.l.*: *s.n.*], 2014. p. 284–287.

PERCIVAL, D. B.; WALDEN, A. T. *et al.* **Spectral analysis for physical applications**. [*S.l.*]: cambridge university press, 1993.

PHUONG, T. M.; LIN, Z.; ALTMAN, R. B. Choosing snps using feature selection. *In*: IEEE. **2005 IEEE Computational Systems Bioinformatics Conference (CSB'05)**. [*S.l.*], 2005. p. 301–309.

PINCUS, S. M.; GLADSTONE, I. M.; EHRENKRANZ, R. A. A regularity statistic for medical data analysis. **Journal of clinical monitoring**, Springer, v. 7, n. 4, p. 335–345, 1991.

POULARIKAS, A. D. **Transforms and applications handbook**. 3ed.. ed. [*S.l.*]: CRC, 2010. 914 p. ISBN 1420066528,978-1420066524.

PRASETIYOWATI, M. I.; MAULIDEVI, N. U.; SURENDRO, K. The speed and accuracy evaluation of random forest performance by selecting features in the transformation data. *In*: **Proceedings of the 2020 The 9th International Conference on Informatics, Environment, Energy and Applications**. New York, NY, USA: Association for Computing Machinery, 2020. (IEEA 2020), p. 125–130. ISBN 9781450376891. Disponível em: https://doi.org/10.1145/3386762.3386768.

QAISAR, S. M.; HUSSAIN, S. F. Effective epileptic seizure detection by using level-crossing eeg sampling sub-bands statistical features selection and machine learning for mobile healthcare. **Computer Methods and Programs in Biomedicine**, Elsevier, v. 203, p. 106034, 2021.

QIAN, B.; RASHEED, K. Hurst exponent and financial market predictability. *In*: PROCEEDINGS OF THE IASTED INTERNATIONAL CONFERENCE CAMBRIDGE, MA. **IASTED conference on Financial Engineering and Applications**. [*S.l.*], 2004. p. 203–209.

RAGHU, S. *et al.* A novel approach for classification of epileptic seizures using matrix determinant. **Expert Systems with Applications**, Elsevier, v. 127, p. 323–341, 2019.

RAMAKRISHNAN, S.; MURUGAVEL, A. Epileptic seizure detection using fuzzy-rules-based sub-band specific features and layered multi-class svm. **Pattern Analysis and Applications**, Springer, v. 22, n. 3, p. 1161–1176, 2019.

RAMALINGAM, A.; KRISHNAN, S. Gaussian mixture modeling using short time fourier transform features for audio fingerprinting. *In*: IEEE. **2005 IEEE International Conference on Multimedia and Expo**. [*S.l.*], 2005. p. 1146–1149.

REFAEILZADEH, P.; TANG, L.; LIU, H. Cross-validation. **Encyclopedia of database systems**, Springer, v. 5, p. 532–538, 2009.

RENNIE, J. *et al.* Tackling the poor assumptions of naive bayes classifiers (pdf). **ICML.**, 2003. Disponível em: https://www.t.ly/lyQ4.

RIAZ, F. *et al.* Emd-based temporal and spectral features for the classification of eeg signals using supervised learning. **IEEE Transactions on Neural Systems and Rehabilitation Engineering**, IEEE, v. 24, n. 1, p. 28–35, 2016.

ROSTAMI, M.; BERAHMAND, K.; FOROUZANDEH, S. A novel community detection based genetic algorithm for feature selection. **arXiv: Learning**, 2020.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. **nature**, Nature Publishing Group, v. 323, n. 6088, p. 533–536, 1986.

SHANNON, C. E. A mathematical theory of communication. **The Bell system technical journal**, Nokia Bell Labs, v. 27, n. 3, p. 379–423, 1948.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). **Biometrika**, JSTOR, v. 52, n. 3/4, p. 591–611, 1965.

SHARMA, M.; BHURANE, A. A.; Rajendra Acharya, U. Mmsfl-owfb: A novel class of orthogonal wavelet filters for epileptic seizure detection. **Knowledge-Based Systems**, v. 160, p. 265–277, 2018. ISSN 0950-7051.

SHIN, H. W. *et al.* Review of epilepsy-etiology, diagnostic evaluation and treatment. **Int J Neurorehabilitation**, v. 1, n. 130, p. 2376–0281, 2014.

SHOEIBI, A. *et al.* A comprehensive comparison of handcrafted features and convolutional autoencoders for epileptic seizures detection in EEG signals. **Expert Systems with Applications**, Elsevier, v. 163, p. 113788, 2021.

SLEPIAN, D.; POLLAK, H. O. Prolate spheroidal wave functions, Fourier analysis and uncertainty—I. **Bell System Technical Journal**, Wiley Online Library, v. 40, n. 1, p. 43–63, 1961.

STUART, S. *et al.* Cortical activity during walking and balance tasks in older adults and in people with parkinson's disease: A structured review. **Maturitas**, v. 113, p. 53–72, 2018. ISSN 0378-5122.

TATUM, W. O. **Handbook of EEG Interpretation**. [*S.l.*]: Demos Medical Pub, 2007. 276 p. ISBN 1933864117,978-1933864112.

THARWAT, A. Classification assessment methods. **Applied Computing and Informatics**, Emerald Publishing Limited, 2020.

THOMSON, D. J. Spectrum estimation and harmonic analysis. **Proceedings of the IEEE**, IEEE, v. 70, n. 9, p. 1055–1096, 1982.

TSIPOURAS, M. G. Spectral information of EEG signals with respect to epilepsy classification. **EURASIP Journal on Advances in Signal Processing**, SpringerOpen, v. 2019, n. 1, p. 1–17, 2019.

TÜRK, Ö.; ÖZERDEM, M. S. Epilepsy detection by using scalogram based convolutional neural network from eeg signals. **Brain sciences**, Multidisciplinary Digital Publishing Institute, v. 9, n. 5, p. 115, 2019.

TZALLAS, A. T.; TSIPOURAS, M. G.; FOTIADIS, D. I. Epileptic seizure detection in eegs using time–frequency analysis. **IEEE transactions on information technology in biomedicine**, ieee, v. 13, n. 5, p. 703–710, 2009.

VAKKURI, A. *et al.* Time-frequency balanced spectral entropy as a measure of anesthetic drug effect in central nervous system during sevoflurane, propofol, and thiopental anesthesia. **Acta Anaesthesiologica Scandinavica**, Wiley Online Library, v. 48, n. 2, p. 145–153, 2004.

VARGAS, D. L. d.; OLIVA, J. T.; TEIXEIRA, M. Uma abordagem baseada em redes neurais artificiais sobre o espectro de potência de eletroencefalogramas para o auxílio médico na classificação de crises epiléticas. *In*: SBC. **Anais do XXI Simpósio Brasileiro de Computação Aplicada à Saúde**. [*S.l.*], 2021. p. 141–152.

WEI, G. *et al.* A novel hybrid feature selection method based on dynamic feature importance. **Applied Soft Computing**, v. 93, p. 106337, 2020. ISSN 1568-4946.

World Health Organization. Epilepsy fact and world health organization and others sheet. **Planned Parenthood Federation of America**, 2014.

World Health Organization. **Epilepsy: a public health imperative**. [*S.l.*]: World Health Organization, 2019.

YANG, Y.; PEDERSEN, J. O. A comparative study on feature selection in text categorization. *In*: NASHVILLE, TN, USA. **Icml**. [*S.l.*], 1997. v. 97, p. 35.

ZHANG, H. The optimality of naive bayes. **Aa**, v. 1, n. 2, p. 3, 2004.

ZHOU, S.-M.; GAN, J. Q.; SEPULVEDA, F. Classifying mental tasks based on features of higher-order statistics from eeg signals in brain–computer interface. **Information Sciences**, v. 178, n. 6, p. 1629–1640, 2008. ISSN 0020-0255.

ZILIAK, S. P values and the search for significance. **Nature Methods**, v. 14, n. 1, p. 3–4, 2017.

ZURLINI, G. *et al.* Highlighting order and disorder in social–ecological landscapes to foster adaptive capacity and sustainability. **Landscape Ecology**, Springer, v. 28, n. 6, p. 1161–1173, 2013.